Lecture Notes in Mechanical Engineering

Peter W. Tse
Joseph Mathew
King Wong
Rocky Lam
C.N. Ko *Editors*

Engineering Asset Management - Systems, Professional Practices and Certification

Proceedings of the 8th World Congress on Engineering Asset Management (WCEAM 2013) & the 3rd International Conference on Utility Management & Safety (ICUMAS)

Springer

Lecture Notes in Mechanical Engineering

About this Series

Lecture Notes in Mechanical Engineering (LNME) publishes the latest developments in Mechanical Engineering—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNME. Also considered for publication are monographs, contributed volumes and lecture notes of exceptionally high quality and interest. Volumes published in LNME embrace all aspects, subfields and new challenges of mechanical engineering. Topics in the series include:

- Engineering Design
- Machinery and Machine Elements
- Mechanical Structures and Stress Analysis
- Automotive Engineering
- Engine Technology
- Aerospace Technology and Astronautics
- Nanotechnology and Microengineering
- Control, Robotics, Mechatronics
- MEMS
- Theoretical and Applied Mechanics
- Dynamical Systems, Control
- Fluid Mechanics
- Engineering Thermodynamics, Heat and Mass Transfer
- Manufacturing
- Precision Engineering, Instrumentation, Measurement
- Materials Engineering
- Tribology and Surface Technology

More information about this series at http://www.springer.com/series/11236

Peter W. Tse · Joseph Mathew
King Wong · Rocky Lam · C.N. Ko
Editors

Engineering Asset Management - Systems, Professional Practices and Certification

Proceedings of the 8th World Congress on Engineering Asset Management (WCEAM 2013) & the 3rd International Conference on Utility Management & Safety (ICUMAS)

Springer

Editors
Peter W. Tse
Department of Systems Engineering and Engineering Management
City University of Hong Kong
Hong Kong
China

Joseph Mathew
Asset Institute
Brisbane
Australia

King Wong
Community and Construction Professionals' Development Centre
Hong Kong
China

Rocky Lam
Engineering Doctorate Society
Hong Kong
China

C.N. Ko
Kam Yuen (Group) International Limited
Hong Kong
China

ISSN 2195-4356
ISBN 978-3-319-09506-6
DOI 10.1007/978-3-319-09507-3

ISSN 2195-4364 (electronic)
ISBN 978-3-319-09507-3 (eBook)

Library of Congress Control Number: 2014955060

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

The 8th World Congress on Engineering Asset Management (WCEAM 2013) was held in conjunction with the 3rd International Conference on Utility Management and Safety (ICUMAS) at the Hong Kong Exhibition and Convention Centre from 30 October to 1 November 2013.

This year's congress was jointly hosted by the International Society of Engineering Asset Management (ISEAM), Department of Systems Engineering and Engineering Management of the City University of Hong Kong (CityU), the Community and Construction Professionals' Development Centre (CCPDC), the Centre for System Informatics and Quality Engineering (CSIE) and the Smart Engineering Asset Management Laboratory (SEAM) of CityU, and the Hong Kong Institute of Utility Specialists (HKIUS).

The joint event attracted 395 registered participants comprising academics, government officials, industrial practitioners and scientists from 35 countries. The event opened and closed with special guests of honour. Mr. Chi-sing Wai, Permanent Secretary (Works) of the Development Bureau, The Hong Kong SAR Government and Professor Way Kuo, President and University Distinguished Professor of the CityU opened the conference, while Mr. Wai-kwok Lo, Member (Functional Constituency—Engineering of the Legislative Council, The Hong Kong SAR government) addressed the audience at the closing ceremony.

Over 190 full papers were submitted and 173 papers were presented during the 3 days of the conference. In addition, the congress also featured 30 esteemed academicians and senior industrialists as the opening, closing and keynote speakers. An industrial forum for Engineering Asset Management practitioners to share their experiences in implementing and obtaining conformance with the British Standards Specification PAS 55 on asset management attracted considerable interest.

In the last decade, Engineering Asset Management (EAM) has continued to evolve into a new professional practice. Through WCEAM we all continuously seek to share the advances in the research and application of this transdiscipline comprising fields such as Management, Finance, Information and Communication Technologies, and Engineering as applied to managing physical assets.

Too much of the developed world's infrastructure and major assets have historically been managed in a somewhat haphazard, reactive, or purely operational way. The absence of a truly strategic, integrated approach to asset management has left many societies now facing challenges of ageing national infrastructure; underinvestment in its maintenance; and a huge gap in addressing the associated climate and sustainability issues.

As its investment in new infrastructure surges, the Asian economies including that of Hong Kong has the serendipitous opportunity to implement planning measures that address asset management issues in the early stages of urban development.

Typically, the term 'asset management' conjures up a range of understanding. From the management of financial (as opposed to physical) assets, to facility management, maintenance management, management of mobile fleets and simply keeping an asset running, asset management has broad application.

In recent years, the study of a fully integrated and highly strategic approach to infrastructure and engineering asset management has emerged as an area of research and development. In this internationally recognised field, asset management is regarded as the process of organising, planning and controlling the acquisition, care, refurbishment and disposal of infrastructure and engineering assets. So much more than just keeping the trains running or coordinating spare parts warehousing and logistics, this emergent field takes a whole-of-life approach to managing all aspects of all assets across an enterprise, with the objective of optimising service delivery potential, minimising cost and risk; and ensuring positive enhancement of natural and social capital over the life of an asset.

Adopting such a view of asset management and implementing an integrated, strategic, approach that utilises technology borne of this emergent research and development would significantly reduce ongoing costs, optimise performance, extend the lifespan and increase the sustainability of the asset base of any enterprise —or nation.

This year's combined congress addresses three themes:

1. Systems, Informatics and Prognosis in Engineering Asset Management
2. Professional Practices and Certification in Engineering Asset Management
3. Underground Assets (utilities) Management in the new Era

Additionally, through the industrial forum the theme "Towards Certification" was also featured as conformance with international standards in asset management is growing in significance around the world. It is likely that when a service provider company wishes to bid for providing public services in the country, the company may need to conform to an EAM standard. Such conformance/certification of EAM has become more than obligatory in a number of countries. Large enterprises are able to hire international EAM consultancy firms to guide them to obtain the required certification. Moreover, in Hong Kong Government bodies, such as Drainage Services Department, private companies such as CLP Power Ltd., MTR Corp., the Hong Kong and China Gas Company Ltd., etc., have already obtained the EAM certification in recent years through PAS 55. However, small and medium

enterprises (SMEs) in Hong Kong cannot afford to hire EAM consultancy companies to guide them. Nearly all enterprises in China do not have much experience in practicing EAM let alone the knowledge to obtain certification. Hence, the congress provided the opportunity to bridge these gaps so that SME owners could learn and share their experience with world experts in EAM as well as certified public utility enterprises. Since EAM is an emerging management technology, the conference provided a convenient and affordable platform in EAM knowledge exchange and experience sharing.

The paper presentations, socials and dinners were of excellent standard. This year's event also provided the International Society of Engineering Asset Management (ISEAM) with the opportunity to present its second Lifetime Achievement award to Professor Emeritus Andrew K.S. Jardine of the University of Toronto in Canada. Professor Jardine is Director of the Centre for Maintenance Optimization and Reliability Engineering. He is also Professor Emeritus in the University's Department of Mechanical and Industrial Engineering. ISEAM Lifetime Achievement awards are awarded for exceptional dedication and contribution to advancing the field of Engineering Asset Management.

All papers published in the proceedings have been refereed by specialist members of an International Review Panel for academic and technical merit.

We gratefully acknowledge the support of Platinum Sponsor ESRI as well as the following sponsors: K.C. Wong Education Foundation, Construction Industry Council—UE Systems Inc, CLP Power Ltd., Hong Kong Electric, The Hong Kong and China Gas Company Limited (Towngas), and Utech Engineering Co. Ltd.

The Chairs are indebted to the Congress Organizing Committee (OC) chaired by Ir Dr. Peter W. Tse, City University of Hong Kong, who was assisted by Dr. C.N. Ko (OC Co-Chair & Chair of Scientific Program Panel, Hong Kong), Mr. Zico Kwok (OC Co-Chair & Chair of Finance Panel, Hong Kong), Professor Lin Ma (OC Co-Chair, Australia), Mr. Paul Tsui (OC Co-Chair & Chair of Promotion & Publicity Panel, Hong Kong), Mr. Raymond Chiu (Chair of Sponsorship & Exhibition Panel & Co-Chair of Scientific Program Panel, Hong Kong), Dr. Rocky Lam (Chair of Technical & Proceedings Panel, Hong Kong) and Mr. Gary Shing-chun Tse (Chair of Sponsorship & Exhibition Panel, Hong Kong).

The next congress in the series, the ninth, will be held in Pretoria from 28 to 31 October 2014 and will be chaired by Professor Joe Amadi-Echendu of the University of Pretoria. For more information, please refer to www.wceam.com.

King Wong
Joseph Mathew
Louis Lock
Kwok-Leung Tsui

International Steering Committee

Kondo H. Adjallah
Basim Al-Najjar
Joe Amadi-Echendu
Veeraragavan Amirthalingam
Ian Barnard
Kerry Brown
Ming Dong
Christos Emmanouilidis
Jin Ji Gao
Marco Garetti
Paulien Herder
Andrew Hess
Stephan Heyns
David Hood
Anthony Hope
Hong Zhong Huang
Benoit Iung
Yi Jia
Jerry Ou Jia-Ruey
Dimitris Kiritsis
Thomas Brett Kirk
Andy Koronios
Helena Kortelainen
Arun Kumar
Gerard Ledwich
Woo-Bang Lee
Jay Lee
Ming Liang
Jayantha Liyanage
Roger Lyon
Lin Ma

Viliam Makis
Joseph Mathew
David Mba
Jun Ni
Khaled Obaia
Jie Pan
Dong Ho Park
Kyoo-Hong Park
Des Pearson
René Pellissier
Bernard Pelei Petlane
Stephen Saladine
Don Sands
Markus Stumptner
Andy Tan
Amy Trappey
Peter W. Tse
Seppo Virtanen
Nalinaksh Vyas
Wen Bin Wang
Margot Weijnen
Roger Willet
Lifeng Xi
Maria Margarita Zelaya-Nunez
Ming Jian Zuo

Order of Presentation

(continued)

(continued)

Contents

An Adaptive Alarm Method for Tool Condition Monitoring Based on Probability Density Functions Estimated with the Parzen Window

Xiaoguang Chen, Guanghua Xu, Fei Liu, Xiang Wan, Qing Zhang and Sicong Zhang

1 Introduction

Tool condition monitoring plays an important role in modern automatic processing for ensuring the processing quality and the machine life [1]. It is important to study the abnormal state and alarm for tool condition monitoring. There are mainly two traditional tool monitoring alarm methods [2]: limit alarm and trend alarm. Limit alarm is that alarm when it is beyond the threshold which is defined statistically by experiences; Trend alarm is that decide whether the faults occurred or will occur by the gradient change of monitoring parameters. As research progresses, many techniques and methods such as Fuzzy Logic (FL) [3], Neural Network (NN) [4] and Support Vector Machine (SVM) [5] have been used for monitoring alarm. They all obtained some effect on improving alarm intelligence and the adaptability of alarm threshold. While such techniques do have following disadvantages: (1)

X. Chen (✉) · G. Xu (✉) · F. Liu (✉) · X. Wan (✉) · Q. Zhang (✉) · S. Zhang (✉)
School of Mechanical Engineering, Xi'an Jiaotong University, 710049 Xi'an, China
e-mail: chenxiaoguang2008@163.com

G. Xu
e-mail: ghxu@mail.xjtu.edu.cn

F. Liu
e-mail: lf0135@gmail.com

X. Wan
e-mail: lianglin@mail.xjtu.edu.cn

Q. Zhang
e-mail: zhangq@mail.xjtu.edu.cn

S. Zhang
e-mail: zhangsicong@mail.xjtu.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_1

working conditions and monitoring parameters cannot be adjusted dynamically by the fixed threshold. (2) The tool operating condition cannot be entirely reflected by the alarm algorithms.

In order to monitor the tool condition and alarm as early as possible, many researchers have investigated different sensors like accelerometer, acoustic emission transducer, current transducer and force sensor [6]. Since the current transducer is non-destructive evaluation, easily installed, non-effective on the normal operation of machine tool, and whose current signals has lower Signal Noise Ratio (SNR), it has been selected for tool condition monitoring throughout this paper.

Research has shown that there exists a good linear relation between the spindle current and the tool wear [7]. The spindle current increases obviously when the cutting tool wears seriously or fails. On the other hand, cutting force is the direct reflection of tool condition, but the force sensor is hard to be installed. However, the change of cutting force leads to the change of the feed current. In general, the spindle current has simple frequency components and is influenced by operating condition greatly, while the feed current has complicated frequency components and is influenced by operating condition slightly. Therefore, the complementary spindle current and feed current is selected in monitoring the tool condition.

In this study, the adaptive alarm method based on probability model with the Parzen window is established. Current signals of the spindle motor and the main feed motor of a CNC are acquired during the tool life. According to their respective features, the amplitude of the spindle current and the root mean square (RMS) of the feed current are extracted. After that, a probability model with the Parzen window is established for data fusion to alarm adaptively.

2 The Probability Model Estimated with the Parzen Window

The Parzen window approach is widely used as a method in probability density estimation. It works properly in small samples and has smooth estimated curve. The Parzen window approach to obtain a non-parametric estimate from a collection of samples is applied as follows [8]. Consider the situation where we have a set of independent samples $X = \{X_1, X_2, \ldots, X_n\}$ with an unknown underlying probability density function $f(X)$. Then the non-parametric estimate of $f(X)$ from X is provided by the function

$$f(X) = \frac{1}{nh^p} \sum_{i=1}^{n} K\left(\frac{X - X_i}{h}\right), \tag{1}$$

where h is the window width coefficient, and p is the sample dimension, and K is a window or kernel function that integrates to unity. In the present work, by using Gaussian window, Eq. (1) is transformed to

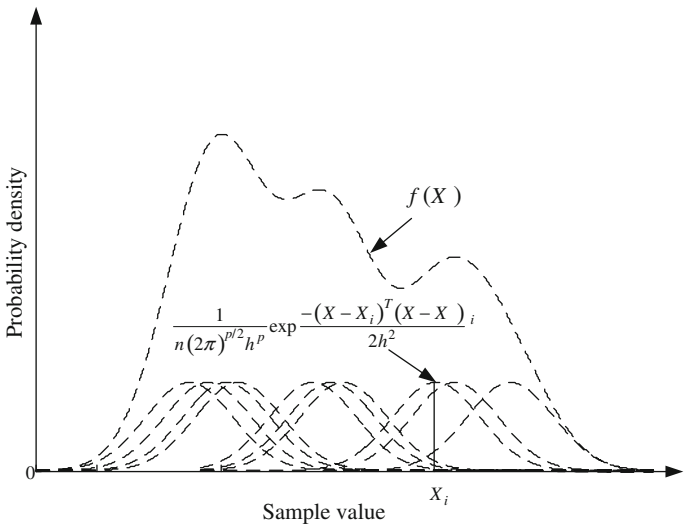


Fig. 1 Parzen window density estimation

$$f(X) = \frac{1}{n(2\pi)^{p/2}h^p}\sum_{i=1}^{n}\exp\frac{-(X-X_i)^T(X-X_i)}{2h^2}. \tag{2}$$

The Parzen window probability model is estimated by using the quantized values of the signals. Its physical interpretation has been shown in Fig. 1, where the total probability density estimation $f(X)$ is the sum of every sample's Gaussian window.

By Eq. (2), the shape of Gaussian window is mainly decided by h. It becomes smoother as h increases, while some details of the density function are buried. When h is small, some details are described enough but the curve can be easily disturbed by random disturbance. Therefore, a proper h is required to balance the above effects. In the present work, h is calculated by

$$h = g * d, \tag{3}$$

where g is an experience constant in [1.1,1.4], and d is the mean minimum distance between samples, given by

$$d = \frac{1}{n}\sum_{i,j=1}^{n}\min d_{ij},\ d_{ij} = \left\|X_i - X_j\right\|,\ (i,j = 1,2,\ldots,n \quad i \neq j). \tag{4}$$

3 Comprehensive Alarm Method for Condition Monitoring

3.1 Alarm Boundary

As shown in Fig. 1, the probability density curve $f(X)$ is close to the real distribution and therefore the boundary is able to describe the monitoring parameters accurately. In the present work, Gaussian window is chosen in Eq. (2), therefore based on the traditional Pauta criterion, the density change caused by X_i that $\|X - X_i\| > 3h$ is just 0.3 %, which can be ignored. Therefore, in the present work, we take the $3h$ neighbourhood of monitoring parameters as the alarm threshold. The $3h$ neighbourhood of the boundary samples is shown in Fig. 2, in which X_l is the low boundary sample and X_u is the upper boundary sample. $X_l - 3h$ is able to represent the low boundary of $f(X)$, and $X_u + 3h$ is able to represent the upper boundary of $f(X)$.

Considering the two-dimensional monitoring parameters, taking every sample value as circle centre and $3h$ as radius, the alarm boundary is formed by every circle. Due to the overlaps of circles, the alarm boundary is the envelope of all circles. Considering the high-dimensional monitoring parameters, the alarm boundary is a complex surface which is the envelope of all $3h$ hyperspheres. The method to determine the alarm boundary is as shown in Fig. 3. After obtaining monitoring parameters, the mean minimum distance would be calculated, then followed by the window width coefficient and the Gaussian window function, we get the alarm boundary.

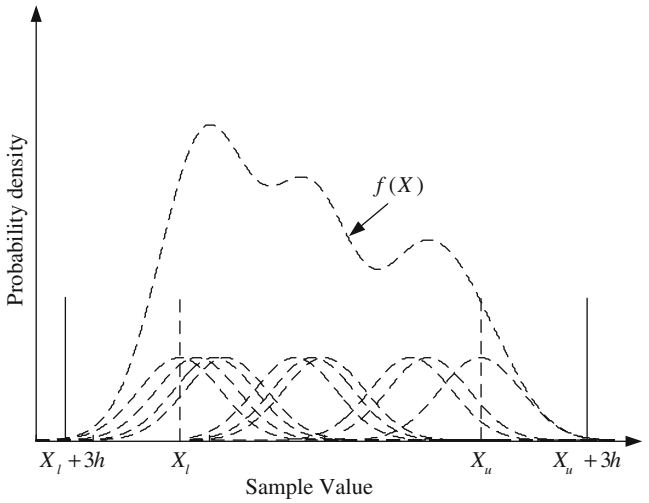


Fig. 2 $3h$ neighbourhood of the boundary samples

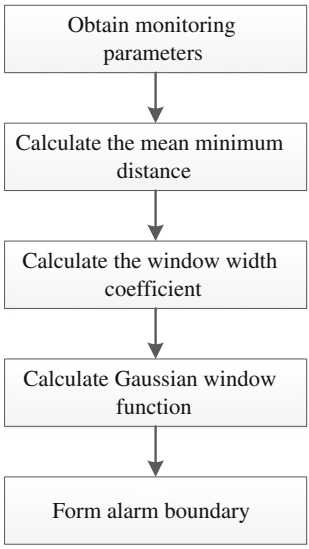


Fig. 3 Method for determining alarm boundary

3.2 Method of Alarm

As the probability model is established by the data acquired in the normal operating condition, alarm occurs when the new monitoring parameter is beyond the alarm boundary. As shown in Fig. 4, the comprehensive alarm method is separated into two aspects based on the distance between the new parameter and the normal

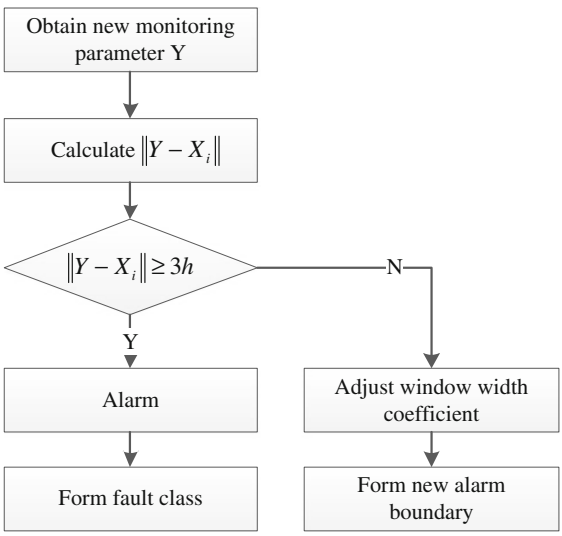


Fig. 4 Comprehensive alarm method

parameters. If the distance $\|Y - X_i\| \geq 3h$, alarm occurs and new fault classes would be built up. If $\|Y - X_i\| < 3h$, the window width coefficient would be adjusted and a new boundary would be formed.

4 Experimental Method

A CNC with cutting tools was studied by Guangzhou CNC Company. Tool life tests were carried out in the state key laboratory under a constant operating condition. The operating condition is shown in Table 1. The current signals were acquired by two commercial current sensors (Fluke i200s). The sample rate is 2 kHz.

In above tests, an optical microscope (KEYENCE VHX-6000) with 1000x magnification was applied for observing the tool wear extent. Figure 5 shows the wear extent in different stages.

5 Signal Processing and Results

Figure 6 depicts the current trends during the whole tool life. It can be seen that the characteristic values did not change much in the early stage since the tool is at the normal wear stage, while the value increases as the wear exacerbates. Also, the value increases sharply when the tool damage occurs.

Table 1 Operating condition

Cutting tool	Rake angle	Relief angle	Work material	spindle speed	Feed speed	Cutting depth
Cemented carbide	6°	8°	Steel 45#	800 rpm	0.5 mm/m	0.5 mm

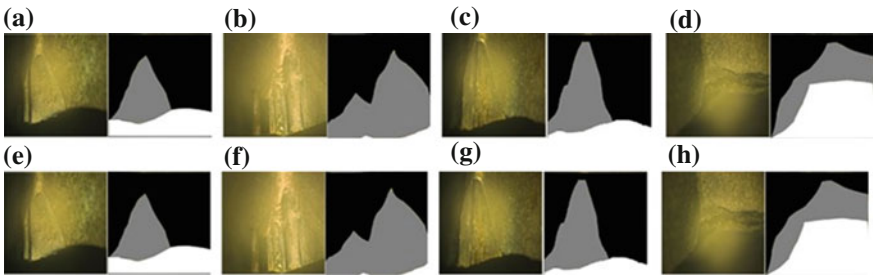


Fig. 5 Images through tool life. **a** New tool. **b** Initial wear. **c** Fast wear. **d** Stable wear. **e** Stable wear. **f** Stable wear. **g** Sharp wear. **h** Breakage

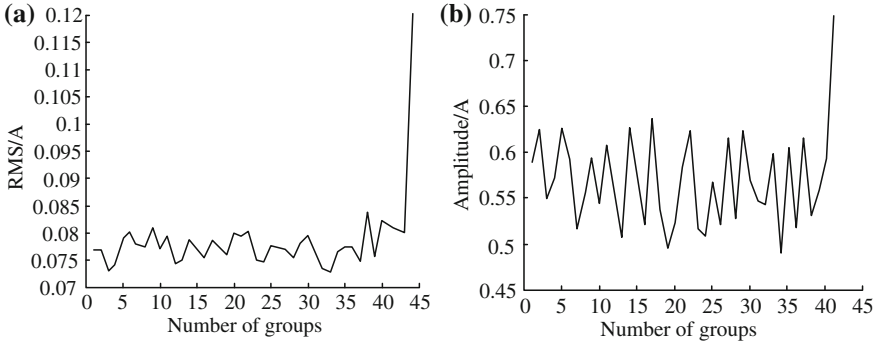
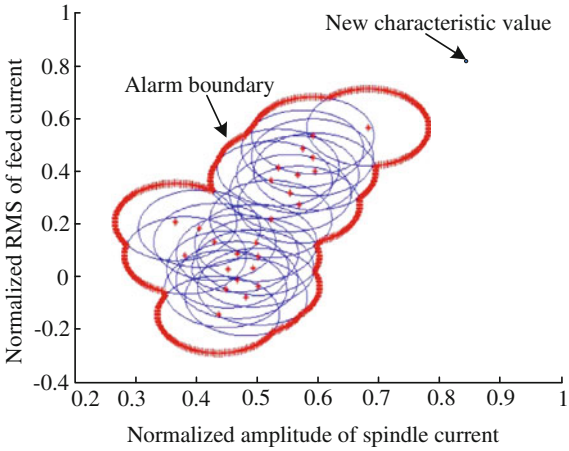


Fig. 6 Current trend through the tool life. **a** RMS of feed current. **b** Amplitude of spindle current

Fig. 7 Result of the alarm method



Due to the large differences of the characteristic values between the feed current and the spindle current, they are first normalized and then analysed by the alarm method. Figure 7 illustrates the result of the alarm method. It shows that the alarm boundary is the envelope. As the new characteristic value is beyond the boundary, this method gives alarm, which means the tool wears seriously or fails and the machine should be stopped. Compared with Fig. 5, the result shown in Fig. 7 matches the fact and proves that this technique alarms adaptively under the occurrence of the cutting tool fracture and is able to meet the factual requirements.

6 Conclusions

In this paper, we proposed an adaptive alarm method based on probability model with the Parzen window by two steps. First, we calculated the window width coefficient with the help of the mean minimum distance between samples. Then it

follows that the probability density function is estimated with the Parzen window and the alarm boundary has been formed. The main conclusions are as follows:

1. The non-contact current sensor could be considered as an attractive quality for tool condition monitoring.
2. Under data-accumulation, the window width coefficient and the alarm boundary are adjusted adaptively to make the alarm more accurate and adaptive.
3. The experimental result proves that this alarm method provides an adaptively and rapidly corresponding alarm when the cutting tool fracture occurs.

References

1. Dimla DE Sr, Lister PM (2000) On-line metal cutting tool condition monitoring. I: force and vibration analyses. Int J Mach Tools Manuf 40(5):739–768
2. Liu G, Hao J, Kuang J (2000) The machine plant accident alarm with trend and overage synthesis method. J Agric Univ Hebei 23(4):89–92
3. Frank PM (1995) Residual evaluation for fault diagnosis based on adaptive fuzzy thresholds. IEE Colloquium (Digest). IEE Press, London pp 1–11
4. Zhang Q, Xu G (2006) Incipient fault diagnosis based on moving probabilistic neural network. J Xi'an Jiaotong Univ 40(9):1036–1040
5. Zhang Q, Xu G, Hua C et al (2009) Self-adaptive alarm method for equipment condition based on one-class support vector machine. J Xi'an Jiaotong Univ 43(1):61–65
6. Jemielniak K, & Arrazola PJ (2008) Application of AE and cutting force signals in tool condition monitoring in micro-milling. CIRP J Manuf Sci Technol 1(2):97–102
7. Zhu K, Wong YS, Hong GS (2009) Wavelet analysis of sensor signals for tool condition monitoring: A review and some new results. Int J Mach Tools Manuf 49(7–8):537–553
8. Rangaraj RM, Wu Y (2010) Screening of knee-joint vibroarthrographic signals using probability density functions estimated with Parzen windows. Biomed Signal Process Control 5(1):53–58

Fault Degradation State Recognition for Planetary Gear Set Based on LVQ Neural Network

Bin Fan, Niaoqing Hu and Zhe Cheng

Abstract In order to ensure the safety and reliable operation of equipment, reduce accidents and economic loss caused by the mechanical fault or failure, prediction and health management (PHM) technology has attracted more and more attention. As the basis and starting point of fault prediction, degradation state recognition is one of the key steps of PHM, which directly affect the reliability of the equipment failure prediction and the selection of corresponding maintenance strategy. As to the degradation state recognition problem of planetary gear set, firstly, select the proper prognosis feature by evaluating various time-frequency features. Secondly, utilize the learning vector quantization neural network to recognize degradation state of planetary gear set. Finally, validate the effectively of presented method with pre-planted chipped fault experiment of planetary gear set. The results show that the proposed algorithm recognizes the multi-level degradation state effectively, and provide a useful reference for subsequent fault prediction.

Keywords Degradation state recognition · Prognosis feature · Neural network · Learning vector quantization

1 Introduction

As an important class of machinery and equipment in weaponry, transportation, electric power, petrochemical, and other fields, mechanical power transmission system is widely used and plays a crucial role in the domain of national defence and

B. Fan (✉) · N. Hu · Z. Cheng
Laboratory of Science and Technology on Integrated Logistics Support, College of Mechatronics and Automation, National University of Defense Technology, Changsha, China
e-mail: fanbin85510@163.com

N. Hu
e-mail: hnq@nudt.edu.cn

Z. Cheng
e-mail: chengzhe@nudt.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_2

economy. With the high-speed improvement of industrial technological and science, mechanical power transmission system is becoming automation, precision, intelligent friendly, its performance and capabilities are gradually improve into perfection. Meanwhile, its structures become more and more complex, and the work environment is usually wretched, the failure or severe fault will lead to disastrous consequences, so the operational safety and reliability go into essential. Therefore, prediction and health management (PHM) technology has been widely concerned in many areas. Find the emergence of incipient fault as early as possible (early fault detection), to remove potential dangers of accidents; identify the current health state of equipment effectively (degradation state recognition), to formulate a reasonable maintenance plan; predict the failure time of equipment (fault prognosis and remaining useful life estimation), to avoid the accident and maximize the equipment effectiveness are three critical goals in PHM. Among them, the degradation state recognition provides a vital link between early fault detection and remaining useful life prediction, it need to be based on the results of former, and itself is prerequisite for latter [1].

Planetary gear set have compact structure, small volume, wide transmission ratio range and high efficiency, is a kind of widely used gear transmission system. Since long-term continuous work under high load and high speed condition, planetary gear set are susceptible to damage and failure, leading to transmission system work abnormally, and even more serious consequences. Therefore, the degradation state recognition for planetary gear system is essential [2, 3]. But due to the complexity of structure, its damage mode and dynamic response is different from the fixed axis gear system, so it is difficult to detect and identify the typical damage in planetary gear set just by some of traditional features. In addition, the difficulty of its dynamics modeling is large, the high accuracy of the model can't be guaranteed.

In contrast, artificial neural network is a simple and effective method. It regards the target system as a black box, relying solely on the characteristics of the input data, by simulating the way biological neural system works to get the desired output, to identify planetary gear set degradation states which is similar to a classification problem. Neural network is composed of large numbers of neurons interconnected, its special processing capability for nonlinear information, can overcome the disadvantages of traditional artificial intelligence methods in mode recognition, voice recognition, unstructured information processing and other nonlinear problems, so that in the nervous expert systems, pattern recognition, intelligent control, prognosis and other fields has been successfully applied. The vector quantization technique was originally evoked by Tuevo Kohonen in the mid 1980s, it introduced a mechanism of competition in neural network, and widespread used for classification and segmentation problems. Learning Vector Quantization (LVQ) network is a typical classification method which applying this technology [4].

In this paper, the degradation state recognition of chipped fault of planetary gear set was solved as a classification problem. Through evaluation of various commonly used features for the planetary gear set fault diagnosis, choose some suitable

features among preferably to constitute input feature vectors, then use LVQ neural network approach to identify the degradation state of object system, and obtain a good results finally.

2 Learning Vector Quantization Neural Networks

2.1 Network Structure

Topologically, LVQ network consists of three layers, the input layer, hidden layer (competitive layer or Kohonen layer) and the output layer, the network architecture is shown in Fig. 1. Network between the input layer and hidden layer is fully connected, and in the hidden layer and output layer connection weights between neurons value is fixed at 1. Before training the network settings in the input layer and the hidden connections between neurons weights, training process, these weights are gradually modified. The output of competitive layer neurons and output layer neurons are all binary. When an input vector is sent to the network, the hidden neuron whose reference vector closest to the input vector will win the competition, thus produce a "1", and the corresponding output neuron will give the class that input vector belonged to. The other hidden neurons are forced to produce a "0". In LVQ network, each output unit has a known class, so it belongs to a kind of supervised competitive neural network.

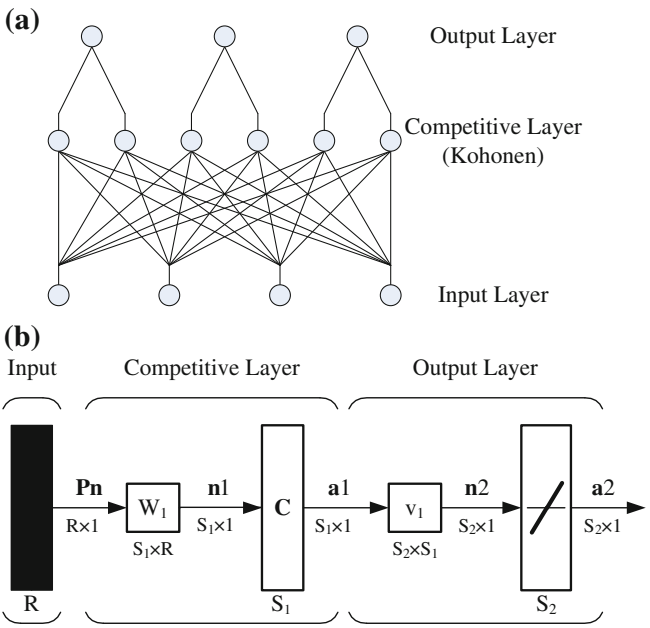


Fig. 1 The LVQ Network. **a** Topology of LVQ, **b** Architecture of LVQ

2.2 Tables, Figures and Pictures

Above all, define the input vector $X = (x_1, x_2,\ldots,x_R)$, wherein, R is the number of neurons in the input layer; the weights matrix between input layer and the competitive layer is $W^1 = (w_1^1, w_2^1, \ldots, w_{s1}^1)$, $w_i^1 = (w_{i1}^1, w_{i2}^1, \ldots, w_{iR}^1)$, w_{ij}^1 represents the weight between ith competitive layer neuron and jth input neuron, $i = 1, 2,\ldots,s_1$, $j = 1, 2,\ldots,R$, and s_1 is the number of competitive layer neurons. The output vectors of competitive layer is $V = (v_1, v_2,\ldots,v_{s1})$, the weights matrix between competitive layer and the output layer is $W^2 = (w_1^2, w_2^2, \ldots, w_{s2}^2)$, where $w_k^2 = (w_{k1}^2, w_{k2}^2, \ldots, w_{ks1}^2)$, w_{kr}^2 represents the weight between rth competitive layer neuron and kth output neuron, $k = 1, 2,\ldots, s_2$, $r = 1, 2,\ldots, s_1$, and s_2 is the number of output layer neurons. It can be classify the input vectors after learning of each neuron in the competitive layer by reference vectors. The class got by competitive layer learning is called subclass, and the class got by output layer learning is called the target class [4].

The learning rule of LVQ neural network combines characteristics of competitive learning and supervised learning. Requires a set of input vectors and their corresponding target classes to train the network, here each target output vector t_j, $j = 1, 2,\ldots,Q$, just has only one component is 1, the other components are all 0. The columns of W^2 represent different subclasses, and rows represent different classes. There is only one component in each column of W^2 is 1, indicating that this subclass belong to the corresponding category. Once defined, W^2 is unchanged.

LVQ network learning is carried out through iteratively update W^1 with improved Kohonen rules for the W^1, specifically speaking, at each iteration, sent an input vector X to the network, and in competitive layer, compute the distances between X and each reference vector, the neuron i^* corresponding to the shortest distance wins competition, so set the i^*th element of output V to 1. Vector Y can be calculated as follows:

$$Y = W^2 V \tag{1}$$

Obviously, there is only one nonzero element in Y. Assumed its number is k^*, that is, X is assigned to class k^*. If X is classified correctly, i.e. $y^* = t_{k^*} = 1$, then move weights of the winner neuron near to X, and update the weights as follows:

$$i^{*w^1}(t+1) = i^{*w^1}(t) + \eta\Big(p(t+1) - i^{*w^1}(t)\Big) \tag{2}$$

If X is classified incorrectly, i.e. $y^* = 1$, but $t_{k^*} = 0$, which means that the wrong hidden layer neurons wins, then move the weights of this neuron away from X, update the weights as follows:

$$i^{*w^1}(t+1) = i^{*w^1}(t) - \eta\Big(p(t+1) - i^{*w^1}(t)\Big) \tag{3}$$

where $\eta \in (0,1)$ is the learning rate, illustrating the adjustment rate of the process for adjusting the weight matrix. $i^{*w^1}(t)$ represents the weights of the i^*th neuron in competitive layer at time t. After network training, each neuron move toward near to vectors of the class it belongs to, and away from vectors belong to the other classes.

3 Feature Choosing and Evaluation

Before the training of LVQ network, it needs to build the input feature vector first. There are many common features used for fault diagnosis or prediction of gearbox, but 21 kinds of them are concerned in this paper [5–7], and are numbered #1–21 in sequence.

Time-domain feature: mean, mean square amplitude, rms, variance, peak, peak to peak, margin factor, crest factor, shape factor, pulse factor, skewness, kurtosis, etc.

Frequency-domain features or **statistical-based features**: FM0, M6A, M8A, NA4, MRS, NF1, NSR, NSR1, spectral kurtosis, etc.

But too many features is a burden of networks, much time is needed to calculate the results, however, the results are not always ideal. In order to solve this problem, a process of feature evaluation and choosing is necessary. By the feature evaluation, choose one or many more suitable features from the feature set above, and construct LVQ network input vector with selected features.

Feature evaluation is assessment of features' sensitivity based on the distance between them. The evaluation rule is: the smaller distance between the feature and other features in the same class, and the larger distance between different classes of the feature, the more sensitive the feature is. Evaluation method can be described as four steps as follows [8].

Step 1: Calculating the average distance of the same class data $d_{i,j}$

$$\begin{gathered} d_{i,j} = \frac{1}{N(N-1)} \sum_{m,n=1}^{N} \left| p_{i,j}(m) - p_{i,j}(n) \right| \\ m,n = 1,2,\ldots,N; m \neq n; i = 1,2,\ldots,K; j = 1,2,\ldots,M \end{gathered} \tag{4}$$

where, N is the number of samples; K is the number of features; M is the number of different classes; $p_{i,j}(m)$, $p_{i,j}(n)$ are sample m and n, i and j represent the number of features and classes, respectively. And then getting the average distance of M classes D_i

$$D_i = \frac{1}{M}\sum_{j=1}^{N} d_{i,j} \tag{5}$$

Step 2: Calculating the average distance between different classes data D_i':

$$q_{i,j} = \frac{1}{N}\sum_{n=1}^{N} p_{i,j}(n) \tag{6}$$

$$D_i' = \frac{1}{M(M-1)}\sum_{u,w=1}^{M} \left|q_{i,u} - q_{i,w}\right| \quad u, w = 1, 2, \ldots, M; u \neq w \tag{7}$$

where, $q_{i,u}$, $q_{i,w}$ are the average value of N samples of the same feature i in class u and w, respectively.

Step 3: Calculating the sensitivity factor, it can be defined as follows:

$$\alpha_i = D_i'/D_i \tag{8}$$

wherein, α_i can reflect the difficulty of classifying M classes use feature i, larger α_i denotes that feature i is more sensitive to different classes, and more suitable for being input vector of network.

Step 4: Sort all features according to α_i from large value to small value, and then choose features for network input vector. First of all, take the first feature as neural network input vector, and then train the network and calculate the test result. Secondly, increase in the number of features one by one, and repeat the calculate process. During certain training epochs, when the training performance, i.e. the Mean Squared normalized Error (MSE) of classification for training samples, decline to 5×10^{-2} or less, it believed that the selected features are enough sensitive to correctly identify M different classes states, and have no further use for other features. Conversely, if training performance is much larger than 10^{-1}, the input vector to add the next feature, until it meets the threshold.

4 Validation

The helicopter transmission fault simulation platform is shown as Fig. 2, it can be used to simulate pitting, chipped, cracks, wearing and other kinds of faults of the planetary gear set system which is the key components of helicopter transmission system. Its design principle is shown as Fig. 3. The drive motor simulates the helicopter engines, and generates driving force; the shaft angle of straight bevel gears 1 and 2, can be used to simulate the spiral bevel gear transmission; #1 and #2

Fig. 2 Helicopter transmission system fault simulation platform

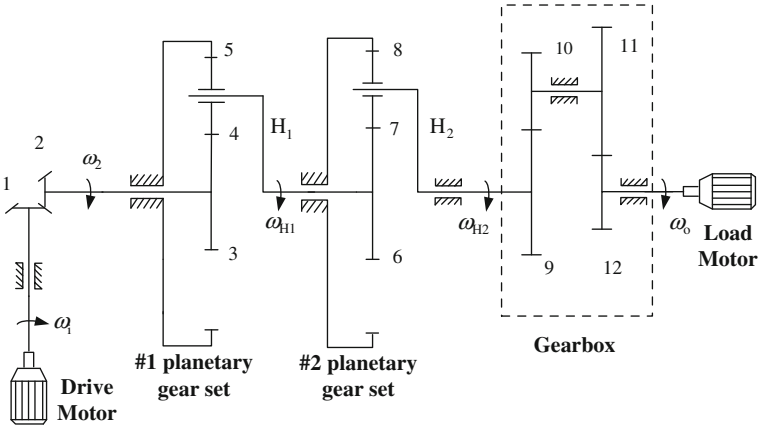


Fig. 3 Illustration of the structure of fault simulation platform

planetary gear sets are used to simulate the helicopter's main reducer; the gearbox can regulate the transmission characteristics between the load motor and the planetary gear set, the load motor also simulates the resistance torque by the rotor shaft. During the experiment, through replacing defective gear, we can study the operating characteristics of transmission system under typical fault condition or evolution of the specific components. Degradation state recognition for the sun gear of planetary gear set is main focus of this paper.

Utilize the simulation platform to take the pre-planted fault experiments. Implant four different severity levels chipped fault in the sun gear of #1 planetary gear set respectively, set the sample frequency is 5 kHz, speed is 1,000 r/min. The acquisition signals include: rotate speed, the horizontal vibration signals (X1) and vertical vibration signals (Y1) of #1 planetary gear set, as well as vertical vibration signals of #2 planetary gear set (Y2). Take 20 samples for every kind of signals, and each sample contains 5,000 points, as shown as Fig. 4.

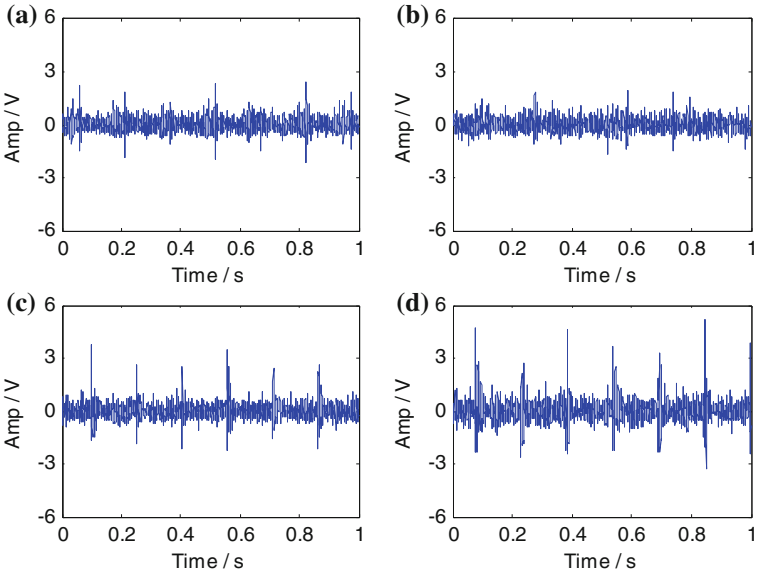


Fig. 4 The different fault level samples of signal X1. **a** level 1, **b** level 2, **c** level 3, **d** level 4

Data processing can be on the procedure as follows:

(1) Extract features from the data obtained before, and get 21 kinds of features to constitute a feature set for each data.
(2) Feature evaluation. Evaluate all features in the feature set of three vibration signal of the different positions respectively, and obtain the result shown as Fig. 5. It is not difficult to find that the sensitive factors of features from different position vibration signals have obvious difference.
(3) Input feature vector construction. According to the above rules, build the feature vector for the three kind of signals respectively. For the signal X1 and Y1, only one feature need to constitute the input vector, which are #18 feature (NF1) and #11 feature(skewness); for the signal Y2, need 7 features to

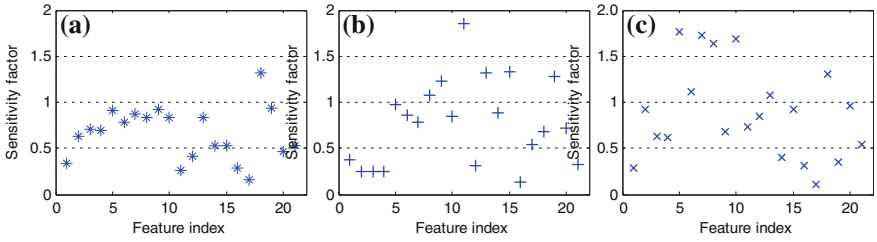


Fig. 5 The results of feature evaluation. **a** X1, **b** Y1, **c** Y2

constitute the input vector, include features #5(peak), #6(peak to peak), #7 (margin factor), #8(crest factor), #10(pulse factor), #13(FM0) and #18(NF1).

(4) Network training. Take 5 samples from every group degradation data samples, total 20 samples as the training data set, and the remaining 60 samples as the validating data. Use neural network toolbox in Matlab software to create an LVQ network and train the network. During the training process, it requires to fix the learning rate η and the number of competitive layer neurons by cross-validation. For signal X1, $s_1 = 11$, $\eta = 0.19$; for signal Y1, $s_1 = 4$, $\eta = 0.04$; for signal Y2, $s_1 = 4$, $\eta = 0.06$.

(5) Degradation state recognition. According to the characteristic of LVQ network, the features of input vector can be taken as the network input directly without normalization. Create an LVQ network with function *newlvq* in neural network toolbox, and after network training, it can be used to identify the state of validation samples. To avoid too long training time, set the training epochs to 300. During the LVQ network training, because of the initialization of competitive layer weights contain some randomicity, the network is a little different from itself after training again, and the recognition results are not exact same. So in order to reduce the fortuity of result, calculating 10 times for every condition, then take the mean of them, and obtain the final recognition result as shown in Table 1.

The results illustrate that LVQ network can effectively identify the four levels fault signals for different positions. Wherein, for signals X1, only a few samples are misidentified as other states in level 3 fault state, the recognition rate reaches 88 %, meanwhile fault samples in level 1, 2 and 4 fault states have been all correctly identified. For signals Y1, fault samples in level 4 fault states have been all correctly identified, there are small amount of misidentified fault samples in other states. Recognition rates for level 1, 3 fault states are around 95 %, and that for level 2 fault states is lower, about 86.7 %. For signals Y2, fault samples in level 1 fault states have been all correctly identified, there are small amount of misidentified fault samples in other states. Recognition rates for level 3, 4 fault states are above 90 %, only for level 2 fault states is about 80 %. Generally, the total recognition rates for three signals are all more than 90, 97 %(X1), 94.3 %(Y1), 92.2 %(Y2) respectively, and demonstrate the effectiveness of LVQ neural network on the degradation state recognition for the experimental data.

Table 1 Identify results of chipped fault based on LVQ network

Fault level	Recognition rate		
	X1 (%)	Y1 (%)	Y2 (%)
Level 1	100.0	96.0	100.0
Level 2	100.0	86.7	80.0
Level 3	88.0	94.7	92.0
Level 4	100.0	100.0	96.7
Total	97.0	94.3	92.2

From the recognition results, it can also be found that, recognition rates are higher for weak fault (level 1) and severe fault (level 4), while other two intermediate fault states are relatively difficult to distinguish, prone to misidentification. In addition, not only feature evaluation results, but also the final recognition results have significant difference for the vibration signals obtained at the different positions. In this paper, the recognition results of signals X1 and Y1 which more closer to the faulty sun gear is better than results of signal Y2.

5 Conclusion

In this paper, we applied the LVQ neural network to identify the degradation state for the sun gear chipped fault of planetary gear set. After the introduction of the structure and characteristic of LVQ network, firstly analyzed the feature evaluation method of input vectors, and then carried out the pre-planted fault experiment with helicopter transmission system simulation platform, acquired enough fault data samples. Finally, trained the network and validated the proposed approach with experiment data. The result illustrated that LVQ network can recognize the degradation state of data samples effectively, and Recognition rate is above 90 %.

Acknowledgment Financial support: This investigation was partly supported by National Natural Science Foundation of China under Grant No. 51075391 and No. 51205401, the Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant No. 20114307110017.

References

1. He ZJ, Chen J, Wang TY, Chu FL (2010) Theories and application of machinery fault diagnostics. Higher Education Press, Beijing
2. Cheng Z, Niaoqing Hu, Gao J (2011) An approach to detect damage quantitatively for planetary gear sets based on physical models. Adv Sci Lett 4(4/5):1695–1701
3. Cheng Z, Niaoqing H, Zhang X (2012) Crack level estimation approach for planetary gearbox based on simulation signal and GRA. J Sound Vib 331(26):5853–5863
4. Sumathi S, Paneerselvam S (2010) Computational intelligence paradigms: theory and applications using MATLAB. CRC Press, New York
5. Samue Paul D, Pines Darryll J (2005) A review of vibration-based techniques for helicopter transmission diagnostics. J Sound Vib 282(1–2):475–508
6. Wu B, Saxena A, Patrick R , Vachtsevanos G (2005) Vibration monitoring for fault diagnosis of helicopter planetary gears. In: Proceedings of the 16th IFAC World congress on disc, Prague, Cesko, 4–8 July 2005
7. Abdulrahman SS, Sharaf-Eldeen YI (2011) A review of gearbox condition monitoring based on vibration analysis techniques diagnostics and prognostics. In: Conference proceedings of the society for experimental mechanics series 8
8. Yang BS, Han T, An JL (2004) ART-KOHONEN neural network for fault diagnosis of rotating machinery. Mech Syst Signal Process 18(3):645–657

A Support System for Selecting the Most Suitable Spare Parts Strategy

Pablo Viveros, Christopher Nikulin, Adolfo Crespo, René Tapia, Raúl Stegmaier, Edward Johns and Fredy Kristjanpoller

Abstract This paper reports an algorithm to create a decision support system for the most suitable spare parts strategy. It is developed based on risk and cost analysis. The algorithm decision support system makers includes information about the possible scenarios in the maintenance field; moreover, this information is useful to support future strategies, helping to decide the number of the spare parts necessary to increase at the maximum process/equipment availability rate. In this task, the algorithm is based on risk analysis considering two main aspects. The first, related to success and failure probabilities is estimated through Poisson distribution; the second one is related to the expected total cost, considering: storage costs, capital tie-up costs, repair costs, purchase order costs (normal and emergency), opportunity costs, and others.

Keywords Spare parts strategy · Risk and cost analysis · Decision making process

1 Introduction

Spare parts strategies have received much attention within the mining industry due to the high cost of equipment. On the one hand, maintenance activities have to support the spare part strategy considering the reliability and availability of the operations. On the other hand, maintenance strategies have to be focused on the maximization of the available resources of the company. Consequently, companies support their maintenance strategies and activities providing different scenario information about requirements, equipment and knowledge of the process [1, 2].

P. Viveros (✉) · C. Nikulin · R. Tapia · R. Stegmaier · E. Johns · F. Kristjanpoller
Universidad Técnica Federíco Santa María, Avda España 1680, Valparaíso, Chile
e-mail: pablo.viveros@usm.cl

A. Crespo
Catedrático de Organización de EmpresasDirector del Dpto. Org. Industrial y Gestión de Empresas IEscuela Técnica Superior de Ingenieros. Universidad de Sevilla, Camino de los Descubrimientos s/n, 41092 Sevilla, Spain

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_3

Therefore, maintenance strategies have to be considered as a systematic approach of measuring and monitoring a set of different variables which include: time, space, resources, production, costs, etc. [3].

Specifically in the field of stocking strategies, many models have been developed by researchers, answering the basic questions: What to stock? Where to stock? How much to stock? Designing the most suitable stock management policy has to take in account several factors such as the characteristics of: demand, criticality, value and specificity of the spare parts. Inventory concepts and approaches, specially designed for the spare parts business, have already been developed for simple repair shops [4], multi-hub systems [5], closed-loop supply chains [6, 7], and multi-echelon supply chains [8–10]. Prasad [11] proposed two different approaches to categorize inventory models: Economic Order Quantity (EOQ) and Material Requirement Planning (MRP). Several case studies have been developed in different fields, including: industrial manufacturing [12, 13], the mining industry [1, 2], the computer industry [14, 15], the electronics industry [16], power generation [17] and in the military [18]. These researchers applied theoretical proposals and concepts to specific settings. However, the complexity of the overall set of variables in order to develop the maintenance strategies causes difficulties in the decision making process. Moreover, strategies to provide a correct decision for spare parts strategy and inventory management still remain as critical issues [19].

Based on aforementioned premise, this paper presents a creative step-by-step algorithm for selecting the most suitable spare parts strategy. Section 2 describes the literature review necessary for the proposal. Section 3 presents a mathematical proposal for calculating the probability of the different scenarios. Section 4, explains an algorithm used to select a spare parts strategy based on the number of spare parts. The final section is dedicated to discussion and conclusion of the proposal.

2 Literature Review

2.1 Risk Analysis

Much research in recent years has focused on psychological and mathematical definitions of risk. On the one hand, psychological (informal) risk definitions proposed by [20], states that risk can be considered as "lack of perceived controllability," "set of possible negative consequences" and "fear of loss". On the other hand, mathematical risk can be considered as the probability of non-controllable events. Some works have proposed measuring risk as a combination of these approaches; e.g. an integral approach of both mathematical and psychological definitions is treated by Suddle and Waarts [21].

The common definition of risk (associated with hazard) is the probability that a hazard will occur and the (usually negative) consequences of that hazard. In essence, it comes down to the following expression (the most frequently used definition in risk analysis) [22].

$$Risk = \sum_{i=1} P_{fi} * C_{fi} \quad (1)$$

The Eq. (1) is the risk equation proposed by Kaplan and Garrick for multiple scenarios (1981) [23], where R is the risk, P_f f is the probability of failure and C_f represents the consequences of the unwanted event.

According to Kaplan and Garrick [24], risk consists of three components; (1) the scenario, (2) the probability of the scenario and (3) the consequences of the scenario [23]. Also suggests that one has to take all hazards into account, which can be accomplished by summing up all possible hazards (scenarios) with their consequences for a certain activity [24].

2.2 Cost Analysis

Much research in recent years has focused on measuring company costs as operational, replacement of equipment, market, etc. There are several costs that allow companies to understand the risks involved in their decisions. Costs are usually separated into direct and indirect costs. Direct costs are related to the production of goods and service, e.g. material, labour, economic value of the component, etc. Indirect costs are not directly related to the object or product, e.g. insurances, taxes, etc. However, the decision not only depends on the main cost of the activities, but also depends on the variables that can affect the occurrence of these activities [25]. These variables allow decision makers to understand the frequency or occurrence of the activities; others allow understanding of the expected costs in the future (e.g. interest rate). Companies have to identify these costs to get a clear understanding of the situation. As a result, the decision making process is more reliable and suitable to formulate the correct strategies.

2.3 Failure Scenarios

Risk evaluation corresponds to the identification of the probable occurrence of specific events or failure and the consequences of these in relation to production or the process. Consequently, many instruments have to be developed to identify this information. Some of these are: different stage tests, control monitoring and information collection.

The different evaluation scenarios identify two different consequences described below (Fig. 1):

- Success scenario: This scenario considers the probability of the equipment or component surviving until the specific evaluation period t before the r-th failure.
- Failure scenario: This scenario considers the probability of the specific r-th failure event within a time period equal to or less than t.

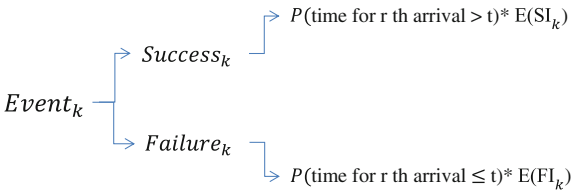


Fig. 1 Description of the failure and success events

where "r" represents the number of failure events to evaluate, "r" has to be greater than zero. The variable "k" represent the number of spare parts available for each scenario, where $0 \leq k \leq r-1$.

2.4 Probability Scenarios

A considerable number of cases use probability scenarios to support decision makers. Probability scenarios attempt to provide quantitative information about possible scenarios. Poisson distribution is a generally accepted methodology for identifying the probability of different scenarios [26]. Consequently an event that causes a specific scenario can be calculated as has been done in Fig. 2a:

According to Fig. 2, λ is the arrival rate of a Poisson process, and k is the number of arrival events (failures/spare parts). Poisson distribution provides the probability of the number of failures over a given amount of time t.

It is knows that $F_r(t) = 1 - R_r(t)$, where $R_r(t)$ is the probability that the number of failures occurring in the time from 0 to t is less than r. For example, decision makers have to consider that the number of events equals the number of failures allowable for the system to continue running during the analysis period t, which should be less than or equal to the number of parts available. It is possible to obtain all the probabilities for the success scenarios. Therefore, the results are probabilistic for discrete scenarios, so that the maximum allowable failures (r−1) and the number of available spare parts are integer numbers.

(a)

$$f(k;t,\lambda) = \frac{(\lambda t)^k * e^{-\lambda k}}{k!} \quad for\, t,\lambda,k \geq 0$$

(b)

$$R_r(t) = \sum_{k=0}^{r-1} \frac{1}{k!} * (\lambda t)^k * e^{-\lambda t}$$

Fig. 2 **a** Poisson distribution; **b** Probability of the number of failures

3 Expected Lifetime Based on Poisson Distribution

Expected Lifetime (ELT) is a usual method for exploring different scenarios based on the life expectancy of the distribution function during periods of time (i.e. 0 to t). ELT has frequently been used in cases with exponential functions. In this case authors provide a modification of the function with 2 parameters such as the Weibull function. Figure 4 provides details of the modification for this paper.

As was represented in Fig. 3a, the Weibull function has made the calculation, where α, β > 0, for the probabilistic analysis of Poisson. The arrival rate of a Poisson Process is calculated as the expectation of the failure rate, as presented in Fig. 3b. On the other hand, to develop a rigorous calculation of life expectancy, it is necessary to work with truncated distribution functions, which must be included in the evaluation period (0 to t) for each stage of evaluation. It is calculated mathematically in Fig. 3c, and graphically represented in Fig. 3d. A discrete form to solve this integral was using the "Riemann Integration" [27], (Fig. 3e). As a result, a complete mathematical development for this case is presented in Fig. 3f, where Γ(r + 1, λt) is an incomplete gamma function, and Γ(r + 1) is a gamma function.

4 Methodology to Select the Spare Part Strategy

This section presents a step-by-step algorithm to support decision makers in their spare parts strategy. The algorithm is described in four main steps starting with the problem context; secondly, the variables analysis; later, the possible scenarios and then, the final decision.

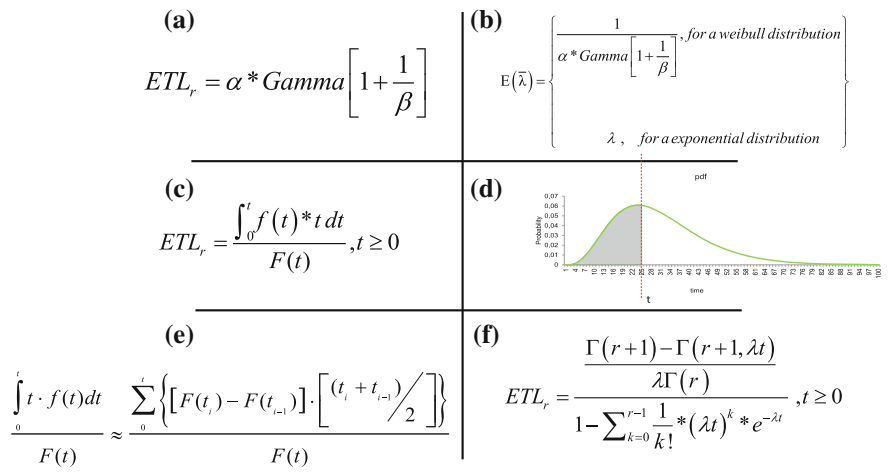


Fig. 3 **a** Expected lifetime; **b** Expected failure rate; **c** Expected failure rate function; **d** Truncation of the probability density function at the time of evaluation t; **e** Riemann integration; **f** Generalized expected failure rate function using Poisson distribution

Step 1: Description of the problem context: In this step it is necessary provide information about the problem context (e.g. equipment or system failure, process downtime, etc.). Decision makers have to describe the system under analysis and/or their parts. It is important identify the different failures within the system; and at the same time hierarchialize criticalities among the elements. This step can be supported by RCA techniques for the identification of failures, and Pareto or Critical Analysis to determine the relevance of these failures. Finally, a part or equipment has to be selected to develop the analysis for different scenarios.

Step 2: Definition of variables and costs: In Step 1 a part or equipment is selected for analysis. In this step the overall set of variables and costs of the problem context have to be identified. This step is performed according to two main aspects:

Step 2.1: Quantify the recurrence of the failure caused by the part or equipment: Identify the Mean Time between Failure (MTBF) and Mean Time to Repair (MTTR); these indicators provide quantitative information about the occurrence of the failure inside the system.

Step 2.2: Identify the costs and variables associated with the failure: The main costs associated with the failure have to be described in this sub-step. Table 1 provides a list of different costs to be considered in the author's proposal.

Step 3: Mapping spare parts scenarios: This step aims to identify the different scenarios according to the information collected from Steps 1 and 2. Two different approaches are necessary to obtain an overall picture of the scenarios. On the one hand, it is necessary to understand the probability of success. On the other hand, it is necessary to identify the scenarios of failure. These two approaches allow companies to compare the consequences of their maintenance strategy from a probabilistic point of view.

Step 3.1: Probability scenarios: This sub-step attempts to provide the probability of each expected scenario. It is mainly supported by Expected Time of

Table 1 Costs associated with the spare parts strategy algorithm

AV	Economic value of the component (USD)
AVC	Interest rate associated to the cost of capital tie-up (% per unit time)
AVS	Interest rate associated to the cost of storing (% per unit time)
OC	Cost of oportunity (%) per hour $(\frac{\text{USD}}{\text{unidad de tiempo}})$
LTE	Logistic time for emergency request (unit time)
CRE	Cost of emergency request of spare parts $(\frac{\text{USD}}{\text{unit}})$
CRN	Cost of normal request of spare parts $(\frac{\text{USD}}{\text{unit}})$
t:	Evaluation period to determine the optical number of spare parts (time)

Life Calculation (ETL). Two different probabilities are obtained for each time (expected scenario of success and failure). Figure 4 shows how to construct the different scenarios according to the number of expected failures and their branches. The number of failures "k" occurring within main scenarios and those failures occurring before or after of expected time "t". The sub-branch appears if the expected time "t" is overcome. Consequently, each scenario has to be considered to obtain the overall picture of the situation.

Step 3.2: Costs scenario: This sub-step attempts to combine the previous sub-step 3.1 with each cost associated with each expected scenario. Therefore, each probability scenario converts into monetary terms where the expected success costs are "E(SI)" and the expected failure costs are "E(FI)".

Step 3.3: Expected success scenario: This sub-step attempts to place a value on the success scenario "E(SI)" according to the costs proposed by author (Table 1). However, the costs are affected by interest rates requiring the costs of each scenario to be modified. Figure 5a represent the expected

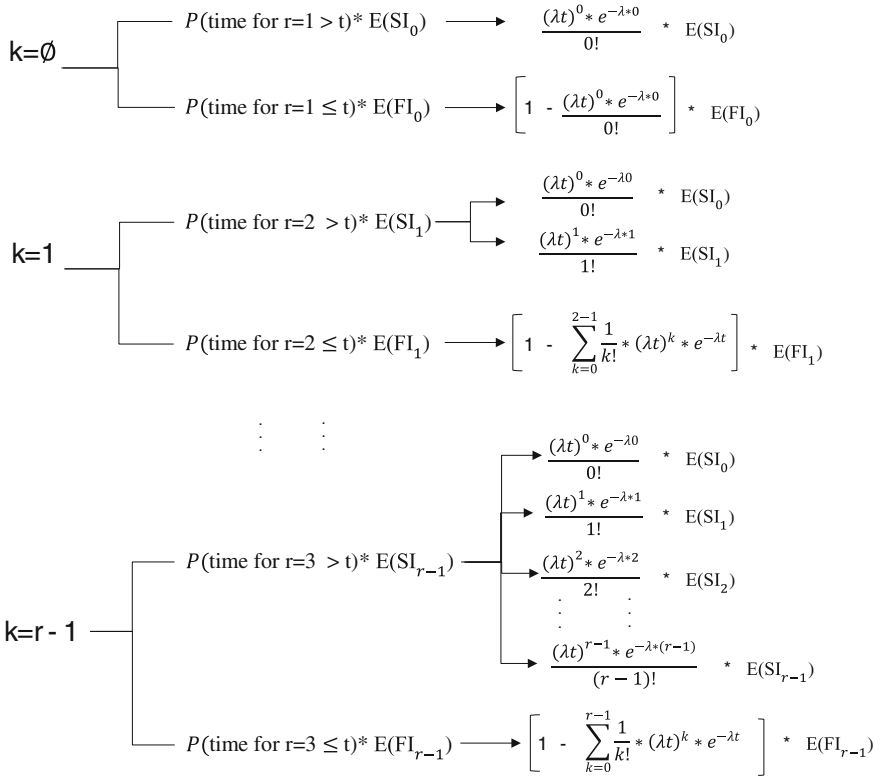


Fig. 4 Event tree for different *r* input, probability of success and failure scenario

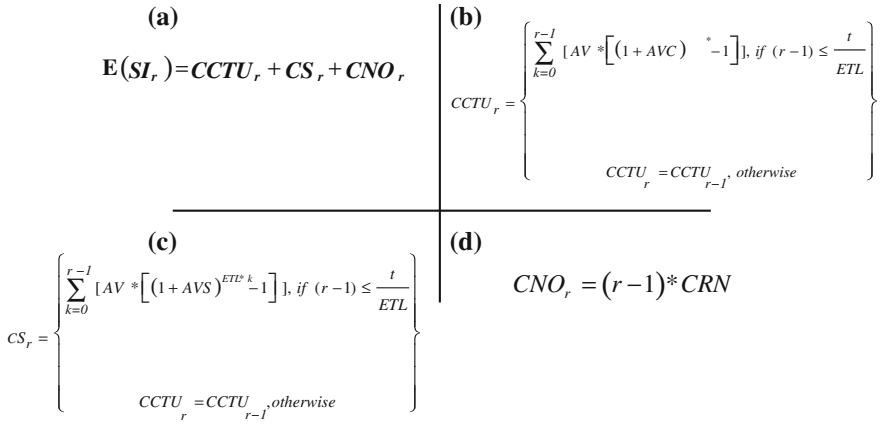


Fig. 5 **a** Expected success scenario; **b** Cost of capital tie-up (CCTU); **c** Cost of store (CS); **d** Cost of normal ordering (CNO)

costs of the success scenario "E(SI)". The modified costs correspond to: Cost of Capital Tie-Up (CCTU), Cost of Store (CS) and Cost of Normal Ordering (CNO). Finally the costs for each r- scenario are calculated as expected success scenarios in Fig. 5.

According to Fig. 5, r represents the r-th failure event causing system failure; therefore (r−1) will be the maximum number of failures allowed during the evaluation time t.

Step 3.4: Expected failure scenario: The expected failure scenario is described using the previous costs plus two additional costs: Cost of Emergency Ordering (CEO) and Cost of Logistic Emergency Ordering (CIL). Figure 6 describes the modified costs for expected failure scenarios.

Step 4: Decision and evaluation of scenarios: From previous steps, it is possible to obtain a clear overview of the different expected scenarios (failure and success). There are many decision variables that can be used in this step to select the best supply chain strategy (e.g. maximum number of spare parts in stock). However the proposed algorithm goes further by emphasizing the author`s recommendation to focus attention on minimum costs strategy in relation to the number of failures according to the different expected scenarios.

The decision making process seeks to determine the optimal number of spare parts (k) that must be available to minimize the risk as much as possible, assuming a finite "t" within the evaluation period, and certainly, sufficient information to develop the evaluation exercises for success and failure scenarios.

(a)

$$\mathbf{E}(FI_r) = CCTU_r + CS_r + CNO_r + CEO_r + CIL_r$$

(b)

$$CEO_r = \begin{cases} CRE * \left[\dfrac{t}{ETL} - (r-1) \right], & if\ (r-1) < \dfrac{t}{ETL} \\ 0, & otherwise \end{cases}$$

(c)

$$CIL_r = \begin{cases} LTE * OC * \left[\dfrac{t}{ETL} - (r-1) \right], & if\ (r-1) < \dfrac{t}{ETL} \\ 0, & otherwise \end{cases}$$

Fig. 6 **a** Expected failure scenario; **b** Cost of emergency ordering (CEO); **c** Cost of logistic emergency ordering (CIL)

The search for a solution is generated by emphasizing the value of r, which directly determines the value of k (k = r−1), where k ≥ 0.

It is necessary to find the policy mix *(r, t),* i.e., optimizing the decision function "*f(r)*" of the evaluation period (interpreted as time provisioning) and the number of spare parts that must be managed in stock (k = r−1) (Fig. 7a). The function "*f(r)*" is related to the Key Performance Indicator (KPI) which allows identification of the minimum costs between two parameters *(r, t)*, thus obtaining overall optimum for making decisions. Consequently, the identification of the best strategy for the number of spare parts corresponds to the $Min\{KPI_{(r,t)}\}$, where r > 1 and t > 0. In principle, it is possible develop a 3D graphs (Fig. 7b), where variables are r, t and $KPI_{(r,t)}$.

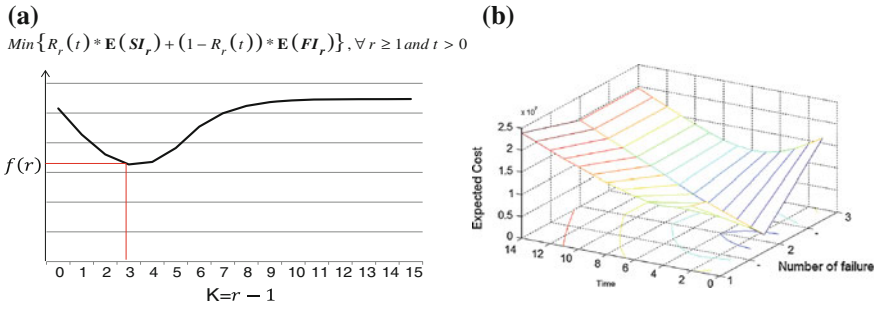


Fig. 7 **a** 2D graph to determine the optimal policy (r, t); **b** 3D graph to determine the overall set of policy r, t and $KPI_{(r,t)}$

5 Discussion

Proposed algorithm attempts to help decision makers identify the best scenarios for spare parts strategy based on reliability and maintainability information, considering the different costs, direct and indirect. However, it is important to add some considerations with respect to the usability of the algorithm. Companies need to identify reliable information for each variable and cost. If the companies don't provide this information the algorithm can be applied on the basis of assumption, but this can cause a negative effect on the accuracy of the final results.

Another point to consider is that the spare parts strategy changes with time. For this reason, it is necessary to provide this kind of analysis in an ongoing way according to the life cycle of the equipment.

6 Conclusion

The acquisition of spare parts is definitely a need that depends on the characteristics of the importance of the equipment, as is the reliability or probability of failure.

Of additional importance are the operating characteristics of the context, the characteristics of the market that supplies the respective parts and also the existing economic variables that influence the decision making process.

The algorithm presented in this paper is a proposal of guidelines for evaluation and calculation methodology to estimate the optimal number of spare parts that must be purchased to meet the needs of the process for a finite period of time. Overall, the conceptual description of the objective of the algorithm is presented, together with the probabilistic mathematical development that supports the model, and of course, the way it should be analyzed and converted into information for decision-making.

The decision is based on the traditional mechanism of risk assessment Risk = P*C, however, the identification, integration and development of the algorithm presented here has a high value of innovation, both in the overall proposal analysis and the algorithm specification.

Evaluation and change within the variables is a useful mechanism of analysis, necessary to understand the effect they have on some input parameters for selected decision, adding to the fact that, generally speaking, decisions are made only considering partial information.

Finally, it is important to note that the research is still in process, specifically in the development of case studies applied in the mining industry in Chile to validate the proposal. In parallel, the authors are developing an evaluation tool to optimize and integrate all partial algorithms (step by step), developing the assessment methodology automatically. The authors consider this algorithm as a starting point to build a software application capable of helping decision makers in selecting a spare parts strategy in this kind of scenario.

References

1. Ghodrati B, Kumar U (2005) Operating environment based spare parts forecasting and logistics: a case study. Int J Logistics Res Appl 8(2):95–105
2. Ghodrati B, Kumar U (2005) Reliability and operating environment based spare parts estimation approach: a case study in Kiruna Mine, Sweden. J Qual Maint Eng 11(2):169–184
3. Claver D, Daoud A-K, Anis C (2009) Integrated spare parts management. In: Handbook of maintenance management and engineering, Springer, London, pp 191–222
4. Scudder GD (1984) Priority scheduling, and spares stocking policies for a repair shop: the multiple failure case. Manag Sci 30(6):739–749
5. Wong H, Cattrysse D, Van Oudheusden D (2005) Stocking decisions for repairable spare parts polling. Int J Prod Econ 93–94:309–317
6. Fleischmann M, Van Nunen JAEE, Gräve B (2003) Integrating closed-loop supply chains and spare-parts management at IBM. Interfaces 33(6):44–56
7. Spengler T, Schröter M (2003) Strategic management of spare parts in closed-loop supply chains—a system dynamics approach. Interfaces 33(6):7–17
8. Diaz A (2003) Modelling approaches to optimise spares in multi-echelon systems. Int J Logistics Res Appl 6(1/2):51–62
9. Kalchschmidt M, Zotteri G, Verganti R (2003) Inventory management in a multi-echelon spare parts supply chain. Int J Prod Econ 81/82 (3):397–413
10. Caglar D, Li C-L, Simchi-Levi D (2004) Two echelon spare parts inventory system subject to a service constraint. IIE Trans 36(7):655–666
11. Prasad S (1993) Classification of inventory model and systems. Int J Prod Econ 34:209–222
12. Bacchetti A, Saccani N (2012) Spare parts classification and demand forecasting for stock control: investigating the gap between research and practice. OMEGA. Int J Manag Sci 40:722–737
13. Bacchetti A, Plebani F, Saccani N, Syntetos AA (2013) Empirically-driven hierarchical classification of stock keeping units. Int J Prod Econ 143(2):263–274
14. Cohen MA, Kamesam PV, Kleindorfer PR, Lee HL, Tekerian A (1990) Optimizer: IBM's multi-echelon inventory system for managing service logistics. Interfaces 20(1):65–82
15. Ashayeri J, Heuts R, Jansen A, Szczerba B (1996) Inventory management of repairable service parts for personal. computers: a case study. Int J Oper Prod Manag 16:74–97
16. Cohen MA, Zheng Y-S, Wang Y (1999) Identifying opportunities for improving Teradyne's service-parts logistics system. Interfaces 29(4):1–18
17. Bailey GJ, Helms MM (2007) MRO inventory reduction—challenges and management: a case study of the Tennessee Valley Authority. Prod Plan Control 18(3):261–270
18. Rustenburg WD, Van Houtum GJ, Zijm WHM (2001) Spare part management at complex technologybased organizations: an agenda for research. Int J Prod Econ 71(1–3):177–193
19. Sang-Chin Y, Zhong-wei D (2004) Criticality evaluation for spare parts initial provisioning. In: Proceedings of reliability and maintainability, 2004 annual symposium—RAMS, pp 507–513
20. Vlek CAJ (1996) A multi-level, multi-stage and multi-attribute perspective on risk assessment, decision-making and risk control. Risk Policy 1:9–31
21. Suddle SI, Waarts PH (2003) The safety of risk or the risk of safety. In: Safety and reliability, van Gelder, Bedford
22. ISO 31000 (2002) Current draft vocabulary for risk management—ISO/IEC guide 73:2009
23. Kaplan S, Garrick BJ (1981) On the quantitative definition of risk. Risk Anal 1:11–27
24. Shahid S (2009) The weighted risk analysis. Safety Sci 47:668–679
25. Jayabalan V, Dipak C (1991) Sequential imperfect preventive maintenance policies for a reliable system. Indus Eng J 20(6):16–21
26. Jardine A, Tsang A (2005) Maintenance, replacement, and reliability: theory and application, CRC Press, USA, p 350
27. Shisha O (1995) The genesis of the generalized Riemann integral. Comput Math Appl 30 (3–6):207–211

A Fusion Approach with Application to Oil Sand Pump Prognostics

Peter W. Tse and Jinfei Hu

Abstract In industrial field, slurry pumps are widely used to transport mixtures of abrasive solids and liquid in wet mineral processing operations. As working under adverse environment, the performances of slurry pumps are often degraded or the pump system even fails unexpectedly. Therefore, significant resources are invested in programs maintenance to avoid unscheduled downtimes and ensure that the required performance of system is maintained at the maximum efficiency. This work is developed from a particular need in oil-mining industry to monitor the health of slurry pumps. In this study, relevance vector machines (RVM) are utilized to predict the remaining useful life (RUL) of field impellers combined with two-summed exponential function. To solve the non-stationary problem emerged in the data, a novel feature extracting process is designed. Finally, one field dataset is applied to evaluate the effectiveness of the proposed prognosis model. The application result shows good performance on degradation trend and remaining useful life prediction of the pump impellers. Hence the model can well solve the problem when to replace the related components before they area absolutely out of service to avoid the sudden downtime.

Keywords Pump impeller · Remaining useful life · Prognosis · Relevance vector machine

P.W. Tse (✉)
The Smart Engineering Asset Management Laboratory (SEAM), The Croucher Optical Non-Destructive Testing and Quality Inspection (CNDT) Laboratory, Department of Systems Engineering and Engineering Management (SEEM), City University of Hong Kong, Hong Kong, China
e-mail: Peter.W.Tse@cityu.edu.hk

J. Hu
Department of Systems Engineering and Engineering Management (SEEM), City University of Hong Kong, Hong Kong, China
e-mail: jinfeihu-c@my.cityu.edu.hk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_4

1 Introduction

Slurry pumps are widely used to move mixtures of abrasive solids and liquid in wet mineral processing operations. These pumps usually experience severe erosive and or corrosive wear even under normal working conditions. Consequently, their performance gets severely comprised with time. At some point, they even start failing unexpectedly. Therefore it becomes necessary to implement a scheduled preventive maintenance programs capable of predicting the degradation trend and estimating the remaining useful life, so as to ensure a safe, economical, and efficient operation of the pump systems in the field.

To the best of authors' knowledge, there are few research efforts that have been devoted to the problem of condition monitoring of slurry pumps; and only a few of work have been reported on the fault diagnosis of the slurry pumps [1–5]. Qu and Zuo proposed a data processing algorithm to clean the data based on support vector classification and random sub-sampling validation [2]. The method yielded good performance on laboratory pump data. Maio et al. proposed an ensemble approach comprising fuzzy C-means and hierarchical trees for assessing the wear status of oil sand pumps [4]. Good diagnosis performance was noticed when the method was evaluated on the basis of data collected from oil sand pump in the field. Note that all the research works cited above have been limited to the problem of fault diagnosis of slurry pumps by using classification methods. In particular, although the problem had received some attention in other contexts [5, 6], few have addressed 'prognostics' which, arguably, is the most important part of condition-based maintenance on slurry pumps. Due to its ability of prior event analysis, prognosis is more effective than diagnostics in assuring zero-downtime performance of machinery.

The work reported in this paper was conducted in response to a particular oil mining need where, as they are prone to sporadic catastrophic breakdowns, the slurry pumps used need enhanced monitoring. In the oil-mining field, in order to lower costs, it is of great practical importance to have available a condition monitoring method capable of determining when a pump should be overhauled or replaced, or how long the useful life of the pump could be. Reliable prediction of the remaining useful life of pumps is likely to yield considerable cost savings and operational safety improvements. In this paper, relevance vector machines (RVM) combined with aggregated exponential functions are utilized to predict the wear degree of field impellers and their remaining useful life. RVM, which was first introduced by Tipping [7], is a data-driven method with Bayesian treatment of support vector machine (SVM), hence it naturally incorporates the prior knowledge compared with SVM. RVM presents a good mechanism on avoiding over-fitting by implementing a prior on the model weights. In this study, the available data or history data of impeller are first used to extract some useful feature(s); then the feature(s) are trained by the RVM-based model to predict the future degradation evolution. The pump remaining useful life can be estimated by extrapolating the gained degradation evolution curve up to the failure threshold.

The remainder of the paper is organized as follows. After the introduction of the basic theory of RVM in Sect. 2, an application of the prediction on the deterioration trend and RUL based on the vibration-based degradation signals is presented for field oil sand pump. In Sect. 4, the results of the application of the prognosis procedure are presented. In Sect. 5, the prognosis performance of the proposed model applied in the real data is analyzed. Finally, we conclude the paper in Sect. 6.

2 Introduction of Relevance Vector Machine

RVM starts with the concept of linear regression models, which are generally utilized to find the parameter vector $\mathbf{w} = \{w_0, w_1, w_2, \ldots, w_N\}$. For a new input $\mathbf{x}$ ($\mathbf{x} \in \Re^N$), the prediction of $\mathbf{z}$ can be obtained according to the following equation:

$$\mathbf{z} = \Phi\mathbf{w} + \varepsilon_n \tag{1}$$

where $\mathbf{\Phi}$ is a $N \times (N+1)$ design matrix, constructed with the ith row vector denoted by $\mathbf{\Phi}_i(x_n) = [1, K(x_n, x_1), K(x_n, x_2), \ldots, K(x_n, x_N)]$; the offset ε_n is an additional noise component of the measurement with mean zero and variance σ^2. In this way, the likelihood of data set can be written as:

$$p(\mathbf{z}|\mathbf{w}, \sigma^2) = (2\pi\sigma^2)^{N-2} \exp\{-\frac{1}{2\sigma^2}\|\mathbf{z} - \Phi\mathbf{w}\|\} \tag{2}$$

In many applications, due to the singularity of the coefficient matrix in Eq. (1), over-fitting problems may arise during the maximum likelihood estimation of parameters in Eq. (2). This could lead to poor prediction performance. To overcome this problem, Tipping proposed to impose some additional constraints on the parameter vector, $\mathbf{w}$.

In the RVM learning process, the parameter vector $\mathbf{w}$ is constrained by putting a zero mean Gaussian prior distribution on the weights, that is

$$p(\mathbf{w}|\alpha) = \prod_{i=1}^{M} N(w_i|0, \alpha_i^{-1}), \tag{3}$$

where α_i is used to describe the inverse variance of each vector w_i, and $\boldsymbol{\alpha}$ denotes as $(\alpha_1, \alpha_2, \ldots, \alpha_M)$. From this formulation, it can be easily seen that there is an individual hyper-parameter α_i associated with each weight so as to control how far each parameter vector is allowed to deviate from zero.

By Bayes' rule, the posterior probability over all the unknown parameters can be expressed as

$$p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)}{p(\mathbf{z})}, \tag{4}$$

where,

$$p(\mathbf{z}) = \int\int\int p(\mathbf{z}|\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)d\mathbf{w}d\boldsymbol{\alpha}d\sigma^2. \tag{5}$$

However, the solution of the posterior $p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{z})$ in Eq. (4) can't be computed directly since the normalizing integral on Eq. (5) is unable to be executed. Instead, we decompose the posterior as:

$$P(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{z}) = P(\mathbf{w}|\boldsymbol{\alpha}, \sigma^2, \mathbf{z})P(\boldsymbol{\alpha}, \sigma^2|\mathbf{z}). \tag{6}$$

According to Bayes' rule, the posterior distribution over weights can be expressed as

$$p(\mathbf{w}|\boldsymbol{\alpha}, \sigma^2, \mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{w}, \sigma^2)p(\mathbf{w}|\boldsymbol{\alpha})}{p(\mathbf{z}|\boldsymbol{\alpha}, \sigma^2)} \sim N(\mathbf{m}, \sum), \tag{7}$$

where the mean $\mathbf{m}$ and covariance $\boldsymbol{\Sigma}$ are

$$\mathbf{m} = \sigma^{-2}\Sigma\Phi^T\mathbf{z}, \tag{8}$$

$$\boldsymbol{\Sigma} = (\mathbf{A} + \sigma^{-2}\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}, \tag{9}$$

where $\mathbf{A} = diag(\boldsymbol{\alpha}) = diag(\alpha_0, \alpha_1, \ldots, \alpha_N)$.

The probability distribution over the training targets can be obtained by integrating the weights to obtain the marginal likelihood for the hyper-parameters:

$$p(\mathbf{z}|\boldsymbol{\alpha}, \sigma^2) = \int p(\mathbf{z}|\mathbf{w}, \sigma^2)p(\mathbf{w}|\boldsymbol{\alpha})d\mathbf{w} \sim N(0, \mathbf{C}), \tag{10}$$

where the covariance matrix is given by $\mathbf{C} = \sigma^{-2}\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{A}^{-1}\boldsymbol{\Phi}^T$. Then the log probability distribution over the training targets is

$$\ln p(\mathbf{z}|\boldsymbol{\alpha}, \sigma^2) = \frac{N}{2}\ln(\sigma^{-2}) - \frac{1}{2}(\sigma^{-2}\mathbf{z}^T\mathbf{z} - \mathbf{m}^T\boldsymbol{\Sigma}^{-1}\mathbf{m}) - \frac{N}{2}\ln(2\pi) + \frac{1}{2}\sum_{i=0}^{N}\ln(\alpha_i). \tag{11}$$

Thus, the estimated value of the parameter weights $\mathbf{w}$ is given by the mean of the posterior distribution in Eq. (7), and the hyper-parameters $\boldsymbol{\alpha}$ and σ^2 can be estimated by maximizing Eq. (11), which is known as the evidence approximation procedure. Further details on the approximation procedure are available at [7].

3 Application to Vibration-Based Degradation Signals from Oil Sand Pump

In this study, a prognostics method is proposed to assess the pump performance degradation and predict the RUL of pump. The schematic diagram of the developed method is depicted in Fig. 1. It mainly involves the following three steps: data acquisition and feature extraction, sparse dataset acquisition by RVM learning process, and model fitting and prediction by extrapolating the fitted model. More details about each step will be given in the following subsections.

3.1 Data Collection and Feature Extraction

Vibration signals using the same sampling frequency rate (51.2 kHz) were obtained from four accelerometers mounted at four different pump locations, viz. Casing Lower (S1), Casing Discharge (S2), Suction Pipe (S3), and Discharge Pipe (S4). Data collection started to be executed when all components inside the pump had just been renewed. It was continued intermittently for around three months with one sampling per hour until the pump's impeller wore out sufficiently to need replacement. Thus, in total, the pump was subjected to 1,007 measurement hours.

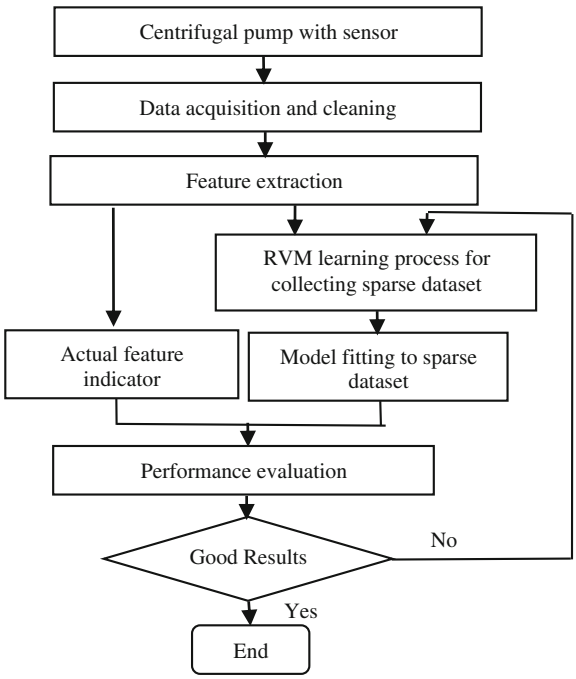


Fig. 1 The schematic diagram of proposed method

The increased vibration levels in certain components of the pump indicated the degradation condition of the pumps, so the vibration signals could be used to monitor pump system health. Data cleaning was done by manually removing outliers exceeding a predefined threshold.

Experience shows that slurry pumps wear mainly because of impeller failure triggered by a decrease in impeller diameter [1]. This suggests that the impeller could be used as the monitored target for pump health assessment and the associated remaining useful life estimation. This research focuses on the pump impeller. Instead of using the rude vibration data directly for health prognosis, a feature extraction procedure was implemented to arrive at good feature(s) pointing to clear progressive degradation of pump impeller.

The vibration data $\mathbf{X}(T, i)$ is firstly standardized in order to scale all the data into the same interval. Next a Fourier transform-based sliding-window averaging technique is employed to obtain averaged FFT amplitude values $\mathbf{Y}(t,f)$ by sliding a window along a sequence of pump-measurement times. Within a narrow spectrum band, 19–40 Hz, the averaged FFT amplitude values $\mathbf{Y}(t,f)$ are then summed up; the summed values are referred as $\mathbf{V}(t)$ and taken into account to substitute for the 'rating frequency' of vane-passing frequency of the monitoring pump. The narrow spectrum band is selected from the overall by manually checking frequency bands of all the pump-measurement times one by one so as make sure that all the situations have been included. The outliers of $\mathbf{V}(t)$ (those exceeding pre-defined threshold) are removed; then the cleaned summed results $\mathbf{V}(k)$, $k = 1, 2, \ldots, K$ are obtained, where K is the reset pump-measurement time ($K \leq 1007 - L + 1$). In the final step, sequential standard deviation values, $STD(j)$, are calculated by augmenting one element out of the cleaned summed results, i.e., $STD(j) = std(V(1), V(2), \ldots, V(j + q - 1))$, where j is referred as the file number index and $j = 1, 2, \ldots, K - q + 1$. $STD(1)$ is calculated from the first q elements which are regarded as the steady stage as impeller deterioration progresses. When compared with the pump damage development demonstrated by the energy evolution based on averaged FFT amplitude as shown in Fig. 2, the standard deviation contains accordant information on pump health condition. Furthermore, it illustrates a monotonically growing trend with damage development along the file number, and thus this is selected as the favorite and effective feature candidate for monitoring the pump health.

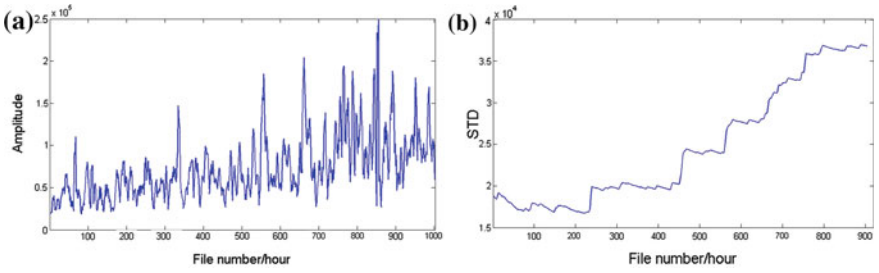


Fig. 2 **a** Energy evolution (T2G1C3). **b** The standard deviation values (T2G1C3)

3.2 RVM Learning Process and Model Fitting

The RVM learning process is performed on the pair of vectors $\{\mathbf{z}, \mathbf{x}\}$, where the input vector $\mathbf{x}$ is constructed from successive inspection file numbers. The target vector $\mathbf{z}$ is constructed by generating the corresponding random numbers which follow the Gaussian distribution with mean values equal to a serial of STD values and variance values equal to a certain pre-defined value. At each inspection file number x_j, $j = 1, \ldots, J$, the target values $\mathbf{z} = \{z_1, z_2, \ldots, z_j\}$ indicating the pump degradation information are assumed to be known up to x_j. For training the RVM model, a Gaussian kernel is employed as the mapping feature space and the value of kernel width is determined via one-dimensional search method 5 to 50 with step length 0.5 with a view to obtaining the optimized RVM training process with the smallest root mean square error (RMS). Thus the hyper-parameters $\mathbf{w}$ and σ^2 in Eq. (1) are determined during the machine learning process. After building the RVM training model, the representative estimators $\mathbf{z}_r^* = \{z_1^*, z_2^*, \ldots, z_r^*\}$ (upon renumbering) whose number is much smaller than that of the training data, are found at the corresponding inspection file numbers $\mathbf{x}_r^* = \{x_1^*, x_2^*, \ldots, x_r^*\}$ (upon renumbering), denoted as a sparse dataset $\{\mathbf{z}_r^*, \mathbf{x}_r^*\}$. An aggregated exponential function is used to fit the pump degradation curve on the basis of the sparse dataset. The future evolution of degradation is predicted by extrapolating the fitted model along the inspection file number and the degradation trajectories are traced up to pre-defined failure threshold; thus simultaneously obtaining the remaining useful life (RUL). In this study, instead of failure thresholds, alert thresholds of pump impellers, beyond which alarms of the pump health are issued and the pump impellers are possible to fail, are set on the basis of our empirical model and pump degradation trend.

4 Results and Analysis of Prognosis Performance

The application of the prognosis procedure on computing the estimated $\hat{RUL}(x_j)$ at the inspection file number x_j of the impeller is hereafter illustrated using two datasets sampled from different positions of the same pump. The dataset is referred to as **T2G1C3**. While verifying the RVM-based model on the basis of empirical data, it is assumed that the equipment may start to fail beyond the maximum degradation level. In these examples, the aggregation of two exponential functions is used to fit the degradation evolution of the impeller, which is derived as follows:

$$z(x) = a \cdot \exp(b \cdot x) + c \cdot \exp(d \cdot x). \tag{12}$$

The corresponding point T_j, at which the alert threshold line and the fitted degradation curve intersect, is derived as

$$T_j = \frac{\log z_F - \log a - \log c}{b + d}, \tag{13}$$

where z_F is the predefined alert threshold value. Hence, the estimated remaining useful life (RUL) is calculated as

$$\hat{RUL}(x_j) = T_j - x_j. \tag{14}$$

Case study: T2G1C3

For case "T2G1C3", the vibration signals were sampled from Suction Pipe. During the feature extraction phase, the sliding window width is selected as 5. The data contained in the first 100 files are taken to represent the steady-state of the impeller. The feature extraction results are plotted in Fig. 3. The alert threshold is set equal to the maximum STD value, and thus the file number at the corresponding intersected point can be easily obtained, i.e., 797. The performance results of the proposed procedure for estimating the impeller RUL by comparing the results obtained by the RVM-based model and the conventional exponential fitting are presented in Fig. 3. Comparisons of the results for dataset T2G1C3 is shown in Table 1. For dataset T2G1C3, the inspection file numbers are selected as $x_j = \{200, 300, 400, 500, 600, 700\}$.

From Table 1, it is clear that the RVM-based model outperforms the exponential fitting. Above all, during the service stage of impeller, the RVM-based model doesn't yield any overestimation of RUL. The empirically observed feature of the proposed method that it does not result in overestimation during the whole working stage is an advantage in practical applications for system maintenance scheduling to avoid cost waste and unexpected downtimes.

The weighted average accuracy of prediction may be calculated using the following formula [8]:

$$\text{Weighted average of accuracy} = \frac{\sum_j \omega_j * \left(1 - \frac{|RUL_A(x_j) - \hat{RUL}(x_j)|}{RUL_A(x_j)}\right) * 100\%}{\sum_j \omega_j}. \tag{15}$$

The weights ω_j are directly proportional to the inspection file number x_j. Note that late predictions are penalized more heavily than early predictions. Hence, the designed performance indicator corresponds with the actual needs and is more trustworthy in practical applications. The results from the weighted accuracy of prediction from the above application are summarized in Table 2.

From the results listed in Table 2, it is evident that the RVM-based model yields much better prediction accuracy compared with that based on exponential fitting. However, it is worth noticing that only one single model (RVM) is adopted to train the prediction system. This might not have been enough to provide a completely robust solution in such a rugged working environment as of the oil sand pumps.

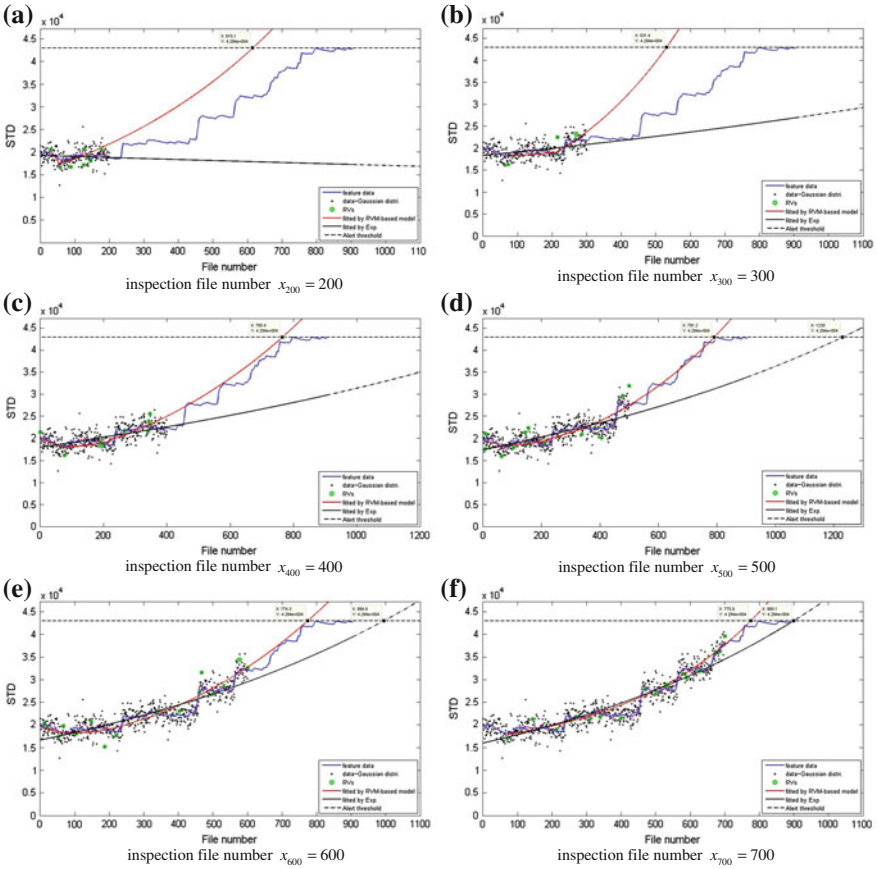


Fig. 3 Comparison prognosis performance by RVM-based model and exponential fitting. **a** inspection file number $x_{200} = 200$. **b** inspection time $x_{300} = 300$. **c** inspection file number $x_{400} = 400$. **d** inspection file number $x_{500} = 500$. **e** inspection file number $x_{600} = 600$. **f** inspection file number $x_{700} = 700$

Table 1 Values of $\hat{RUL}(x_j)$ by RVM-based model and exponential fitting

Inspection file number (x_j)	Actual $RUL_A(x_j)$	$\hat{RUL}(x_j)$ by RVM-based model	$\hat{RUL}(x_j)$ by Exponential fitting
200	597	415	>>1,200
300	497	231	>900
400	397	365	753
500	297	291	730
600	197	174	395
700	97	76	200

Table 2 The weighted average accuracy of prediction for pump impeller

RVM-based model (%)	Exponential fitting (%)
82.02	1.53

5 Conclusions

This paper has presented a model combining relevance vector machines (RVM) and aggregated exponential function that can be used for pump impeller prognosis and the estimation of the pump's remaining useful life (RUL). The data used in the case study are all sampled from the field—a pump in actual use in an oil-mining industry. To solve the non-stationary problem emerging from the vibration data, a novel feature extracting process is proposed to arrive at a feature varying monotonically with damage development of pump impellers. The designed procedure is capable of treating degradation signals for RUL estimation and presents better performance compared with that based on standalone exponential fitting. However, there is still space left for improvement. As our future work, more research effort will be devoted to design novel ensemble prognosis models which can balance out the errors of single models, and further improve the prediction accuracy and robustness.

Acknowledgment The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 122011) and a grant from City University of Hong Kong (Project No. 7008187).

References

1. Zhao X., Hu Q., Lei Y, Zuo M (2010) Vibration-based fault diagnosis of slurry pump impellers using neighbourhood rough set models. In: Proceeding IMechE, Part C: J Mech Eng Sci 224 (4):995–1006. http://pic.sagepub.com/content/224/4/995.short
2. Qu J, Zuo M (2010) Support vector machine based data processing algorithm for wear degree classification of slurry pump systems. Measurement 43(6):781–791. http://www.sciencedirect.com/science/article/pii/S0263224110000552
3. Hancock K, Zhang Q (2006) A hybrid approach to hydraulic vane pump condition monitoring and fault detection. Transactions of American Society of Agricultural Engineers 49(4):1203–1211. http://age-web.age.uiuc.edu/faculty/qzhang/Publications/2006TransASAE49(4)Monette.pdf
4. Di Maio F, Hu J, Tse P, Pecht M, Tsui K, Zio E (2012) Ensemble-approaches for clustering health status of oil sand pumps. Expert Syst Appl 39(5):4847–4859. http://www.sciencedirect.com/science/article/pii/S095741741101493X
5. Shen Z, He Z, Chen X, Sun C, Liu Z (2012) A monotonic degradation assessment index of rolling bearings using fuzzy support vector data description and running time. Sensors 12:10109–10135. http://www.mdpi.com/1424-8220/12/8/10109/pdf
6. Wang D, Tse P, Guo W, Miao Q (2011) Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis. Meas Sci Technol 22 (2):1–14. http://iopscience.iop.org/0957-0233/22/2/025102

7. Tipping M (2001) Sparse Bayesian learning and the relevance vector machine. J Mach Learn Res 1:211–244. Retrieved from Sparse Bayesian learning and the relevance vector machine
8. Caesarendra W, Widodo A, Yang B-S (2010) Application of relevance vector machine and logistic regression for machine degradation assessment. Mech Syst Signal Process 24(4): 1161–1171. http://www.sciencedirect.com/science/article/pii/S0888327009003239

Diagnosis of Air-Conditioner by Using Its Dynamic Property

Tadao Kawai and Seiya Kushizaki

Abstract To avoid failures of an air-conditioner, we need suitable condition monitoring technique. Unfortunately, an air-conditioner has many parameters in it, i.e., inlet temperature, outlet temperature of a condenser or an evaporator, pressure of each component and input power for a compressor and so on. Furthermore, it is very difficult to identify failures in an air-conditioner because of small difference of parameters between normal and abnormal condition. In this paper, we proposed diagnosis technique by using support vector machine (SVM) with identified system parameters by ARX model. Because dynamic property of an air-conditioner identified by ARX model did not depend on operation condition, i.e., room temperature, outside temperature, our proposed technique did not need a lot of data. Finally, our proposed technique well identified the clogging in heat exchanger very well.

Keywords Air-conditioner · Diagnosis · ARX model · SVM

1 Introduction

In these days, air-conditioners are built in almost all buildings to keep temperature and humidity at a suitable level. For Japanese summer is oppressively hot and humid, an air-conditioner is one of the most important equipment for us to do good job and keep comfortable life. An air-conditioner works all day long for business use and long time for personal use. So that, if an air-conditioner goes down, serious situation will be caused especially in business use. Failures occurred in an air-conditioner include trouble in an electric circuit, a cutting of electric wires, a breakdown of motor, a clogging of a heat exchanger and so on. To deal with these

T. Kawai (✉)
Osaka City University, 3-3-138 Sugimoto-Cho, Sumiyoshi-Ku, Osaka, Japan
e-mail: kawai@mech.eng.osaka-cu.ac.jp

S. Kushizaki
Mieden System Solution Co. LTD, Osaka, Japan

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_5

troubles, break down maintenance (BDM) is carried our for personal use air-conditioner and time based maintenance (TBM) is carried out for business use in many cases. Although a self monitoring system is built in an advanced air-conditioner, this system monitors only an electric circuit. In spite of many researchers focused on condition based maintenance (CBM) [1–5], effective monitoring technique is not developed.

Authors focused on a clogging in a heat exchanger being very difficult to detect. We applied support vector machine (SVM) [6, 7] to classified a clogging in a heat exchanger from normal condition using steady-state data [8]. Although proposed technique worked well, we needed a lot of data for learning and had to exclude transitional data caused by start-up of compressor.

In this paper, we estimated dynamic property of an air-conditioner using ARX model with transitional data cause by a compressor and so on and classified condition of an air-conditioner by its property using SVM. Dynamic property did not change in a wide range of temperature and changed sensibly to a clogging of heat exchanger. We checked performance of our proposed technique with two types of kernel function and several parameters.

2 Diagnosis Technique by Using Its Dynamic Property

Figure 1 illustrates a proposed diagnosis technique.

[Step 1: Learning]

Dynamic properties of air-conditioner are identified as ARX model by using transitional data. Then parameters of ARX model are classified by SVM. This procedure is carried out for normal condition and condition with a clogging in an outlet and an inlet of heat exchanger.

[Step 2: Diagnosis]

Using data obtained with unknown condition, dynamic properties are identified as same ARM model. After identified parameters are classified by above SVM, we can detect a clogging.

In our research, two types of kernel functions are tested.

(1) Gaussian kernel

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{\sigma^2}\right)$$

(2) Polynomial kernel

$$K(x, x') = (x^T x' + 1)^d$$

Advantages of proposed technique are listed as follows.

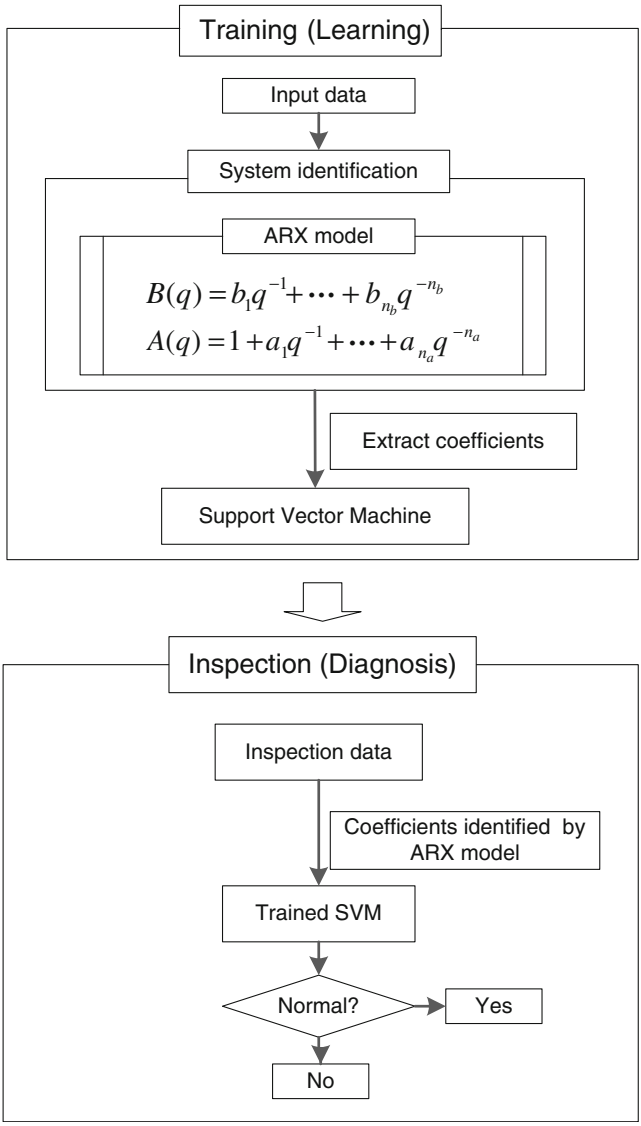


Fig. 1 Procedure of proposed technique

(a) Dynamic properties of an air-conditioner are independent on room temperature or outdoor temperature. This means only transient data caused by a compressor is required for proposed technique.
(b) In general, a failure strongly affects transient data. So that, proposed technique is sensitive to a clogging. In steady-state condition, a clogging changes outlet temperature a little. It is very difficult to monitor this little change from changes of outdoor or room temperature.

3 Verification of Proposed Technique

3.1 Verification by Simulation

3.1.1 Simulation Model

Among many useful and powerful modeling techniques, we used commercial package "Dymola with air conditioning library" of modelica language [9] for building physical model. Figure 2 illustrates a brief image of model.

3.1.2 Simulation Condition

Failure in air-conditioner includes a snapping of a wire or an electric circuit, a clogging in heat exchanger and so on. Among these failures, a snapping of a wire or an electric circuit is easy to detect and a diagnosis system is already built in many commercial products. In this paper we focused on a clogging in heart exchanger (Exp1) and reduction of heat transfer coefficient of heat exchanger (Exp2). In the case of Exp1, dust accumulated between fins in heat exchanger disturbs air flow to prevent heat transfer. In the case of Exp2, accretion on fin prevents heat transfer from fin to air without disturbing air flow.

Table 1 lists effects of above failure estimated by simulation. In the table, "ref" or "air" designates refrigerant and air, "in" and "out" shows an inlet and an outlet.

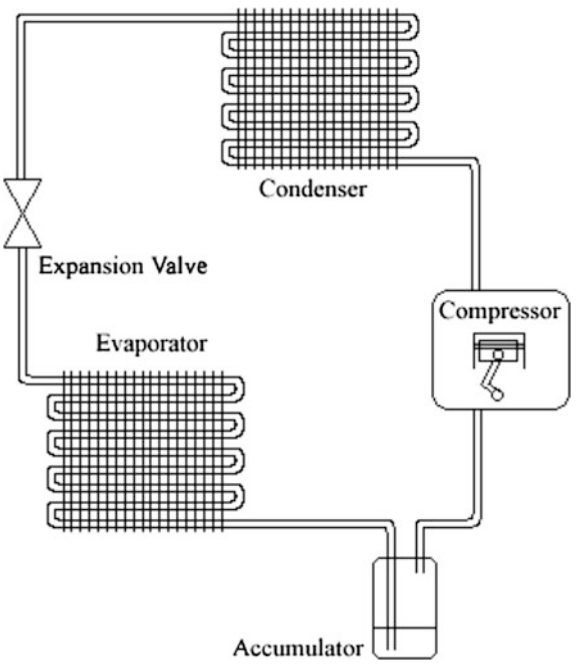


Fig. 2 Simulation model of air-conditioner

While up-arrow means an increase of temperature or pressure, down-arrow means a decrease of temperature or pressure and "-" means no change. Because we neglect compressibility of air, effect of pressure of air is omitted in our consideration and marked as "*". For example, in the case of Exp1 applied for condenser, temperature and pressure of refrigerant at inlet and outlet of condenser and evaporator rise. Furthermore, temperature of air rises at an outlet.

Table 2 shows simulation condition for above mentioned two type of failures. In the table, a column designated by "dynamic behavior" means which component start to work. Here a component is a fan of a condenser, a fan of evaporator or a compressor. A row designated by "state" means what kind of state a system is in, i.e., normal, clogging at condenser or clogging at evaporator. For example, in the case of ①, we simulated without clogging and got dynamic data at a timing when a fan of a condenser started to work. We carried out simulations with fifteen types of condition.

3.1.3 Learning Condition

In our previous report [8], we selected suitable parameters to monitor the condition of an air-conditioner by experiment. Following parameters were essential parameters for monitoring.

Table 1 Effect of clogging on temperature and pressure

			Condenser				Evaporator			
			ref		air		ref		air	
			in	out	in	out	in	out	in	out
Condenser	Exp1	T [K]	↑	↑	–	↑	↑	↑	–	↑
		P [MPa]	↑	↑	*	*	↑	↑	*	*
	Exp2	T [K]	↑	↑	–	↓	↑	↑	–	↑
		P [MPa]	↑	↑	*	*	↑	↑	*	*
Evaporator	Exp1	T [K]	↓	↓	–	↓	↓	↓	–	↓
		P [MPa]	↓	↓	*	*	↓	↓	*	*
	Exp2	T [K]	↓	↓	–	↓	↓	↓	–	↑
		P [MPa]	↓	↓	*	*	↓	↓	*	*

Table 2 Simulation condition

dynamic behavior \ state	Normal	Clog at Condenser		Clog at Evaporator	
Condenser	①	④	⑦	⑩	⑬
Evaporator	②	⑤	⑧	⑪	⑭
Compressor	③	⑥	⑨	⑫	⑮

Fc (rps)	Rotational speed of fan at condenser
Fe (rps)	Rotational speed of fan at evaporator
W (W)	Compressor power
ΔTc (K)	Difference of temperature of refrigerant between inlet and outlet of condenser
ΔTe (K)	Difference of temperature of refrigerant between inlet and outlet of evaporator
ΔTcair (K)	Difference of temperature of air between inlet and outlet of condenser
Teairin (K)	Temperature of air at inlet of evaporator
Teairout (K)	Temperature of air at outlet of evaporator

To estimate parameters of ARX model, we supposed that temperature of air at an outlet of a evaporator as "output" and other parameters as "input".

3.1.4 Result of Diagnosis

Tables 3 and 4 list diagnosis result estimated by SVM with Gaussian kernel and polynomial kernel respectively.

Notation in these tables is as follows.

n	Normal
c-1	Clogging at a condenser
c-2	Drop in heat transfer at condenser

Table 3 Classification result of failure with Gaussian kernel

	data	σ=0.1	σ=0.5	σ=1	σ=2	σ=3	σ=5
Condenser	n	c	c	c	c	c	c
	c-1	c	c	c	c	c	c
	c-2	c	c	c	c	c	c
	e-1	e	e	e	e	e	e
	e-2	e	e	e	e	e	e
Evaporator	n	c	c	c	c	c	c
	c-1	c	c	c	c	c	c
	c-2	n	c	c	c	c	c
	e-1	n	n	n	n	n	n
	e-2	e	c	c	e	e	c
Compressor	n	c	c	c	n	n	n
	c-1	c	c	c	c	c	c
	c-2	c	c	c	c	c	c
	e-1	e	e	e	e	e	e
	e-2	e	e	e	e	e	e

Table 4 Classification result of failure with polynomial kernel

	data	d =1	d =2	d =3	d =4
Condenser	n	e	e	e	e
	c-1	n	n	n	n
	c-2	c	c	c	c
	e-1	e	e	e	n
	e-2	e	e	e	e
Evaporator	n	c	c	c	c
	c-1	c	c	c	c
	c-2	n	n	c	c
	e-1	n	n	n	n
	e-2	e	e	e	e
Compressor	n	c	c	c	n
	c-1	c	c	c	c
	c-2	c	c	c	c
	e-1	e	e	e	e
	e-2	e	e	e	e

e-1 Clogging at an evaporator
e-2 Drop in heat transfer at an evaporator

Transient data at start-up of a condenser fan, an evaporator fan and a compressor with conditions of “n”, “c-1”, “c-2”,“e-1” and “e-2” were used for diagnosis. In columns from σ = 0.1 to 5 in the Table 3, diagnosed result is listed as “n”, “c” and “e”. Here, “n” means normal, “c” means failure at a condenser and “e” means failure at an evaporator. Moreover, in columns from d = 1 to 4 in the Table 4, same designation is used for a diagnosed result. We draw the cells diagonal line to show a wrong diagnosed result. These tables demonstrate well diagnosed result using dynamic property of a compressor, especially with large value of kernel function.

3.2 Verification by Experiment

3.2.1 Experimental Setup

Figure 3 is a target system used for experiment. An air-conditioner is made by Daikin Industries, Ltd. (Cooling power is 2.8 KW, type of compressor is 1YC23AAXDA, type of room fan is QCL9362 M, type of outside fan is PZ370).

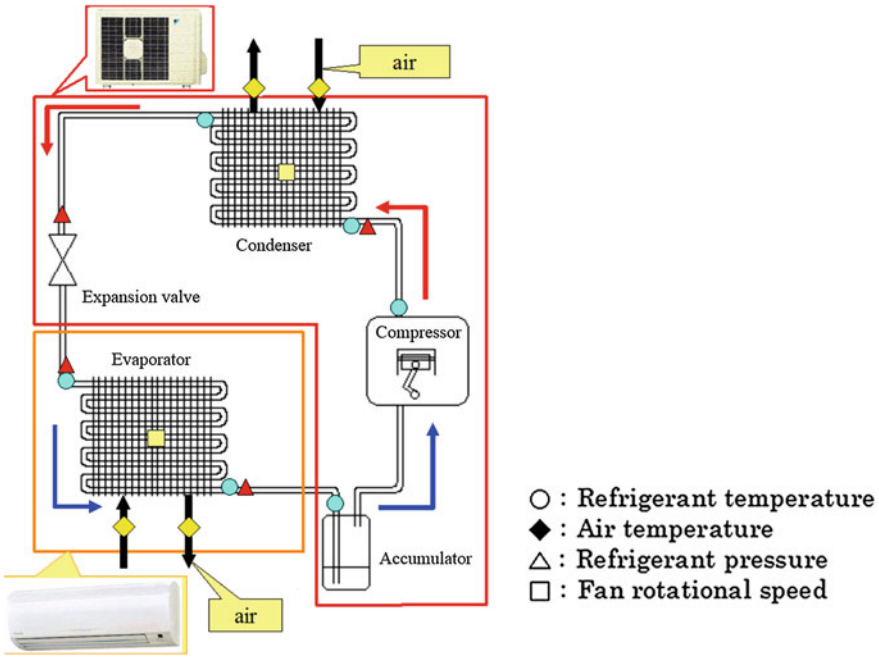


Fig. 3 Experimental setup

Refrigerant used in this air-conditioner is R410A. In the figure, measuring points are designated by marks ○, ♦, Δ and □. Power to a compressor was estimated by current to a compressor.

In the experiment, we covered an indoor unit, i.e., an evaporator, and an outdoor unit, i.e., a condenser, by net or cloth. Net reduced air flow through an indoor unit by 12 %. Cloth reduced air flow through an outdoor unit by 25 %.

3.2.2 Diagnosed Result

Experiments were carried out at points shown in Fig. 4 to get transitional data. In Fig. 4, vertical axis shows a range of outside temperature, i.e.; inlet temperature of a condenser, and horizontal axis shows a range of room temperature, i.e.; inlet temperature of an evaporator. We carried out experiments six times with normal condition, eleven times with covered condenser and seven times with covered evaporator. For learning data, we selected four normal data, seven covered condenser data and four covered evaporator data randomly. Other data were used for diagnosis.

Diagnosed result are illustrated in Figs. 5 and 6. Figure 5a and b show result with Gaussian kernel, and Fig. 6a and b show a result with polynomial kernel. These figures show a ratio of a diagnosed result to given condition. In the case of given

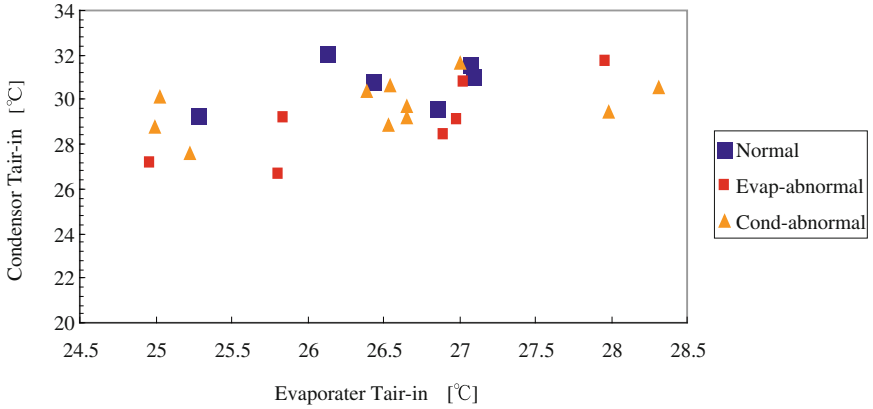


Fig. 4 Selection of dataset for training and diagnosis by SVM

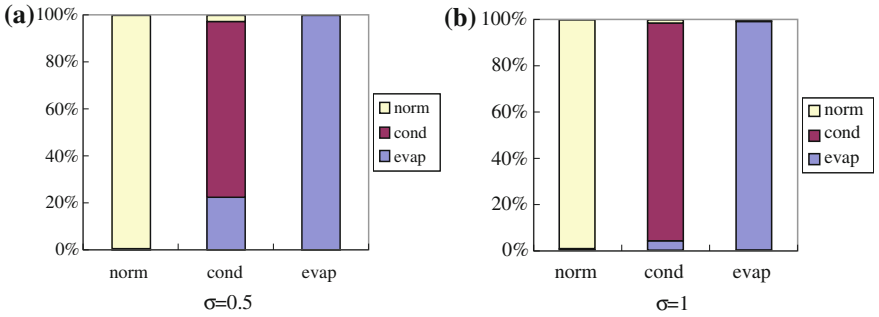


Fig. 5 Classification result by Gaussian kernel. **a** σ = 0.5, **b** σ = 1

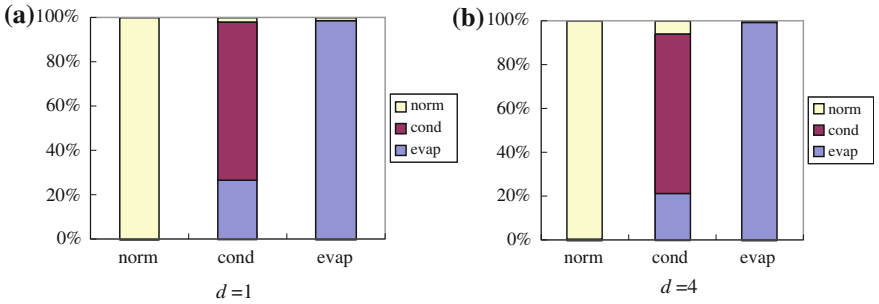


Fig. 6 Classification result by polynomial kernel. **a** d = 1, **b** d = 4

condition of a covered condenser shown in Fig. 5b as a middle bar, estimated percentage of normal, covered condenser and covered evaporator are 1.5, 94 and 4.5 % respectively. These results show that normal and covered evaporator are well identified in all cases. On the contrary, in the case of a covered condenser, estimated result is a little bit worse.

4 Conclusion

In this paper, we proposed diagnosis technique to detect failure in air-conditioner by using ARX model and SVM. We carried out simulation and experiment to estimate an ability of our technique and got following results.

(1) We proposed a technique to estimate dynamic property of an air-conditioner using ARX model with transitional data and diagnose condition of an air-conditioner by SVM. The proposed technique was independent on operating condition of an air-conditioner and required not so many data.
(2) In our experiments, we identified a normal condition and a covered evaporator almost 100 % in all cases and identified a covered condenser about 94 % at most.

References

1. Ghiaus C (1999) Fault diagnosis of air conditioning systems based on qualitative bond graph. Energy Build 30(3):221
2. Wang, S, Wang JB (1999) Law-based sensor fault diagnosis and validation for building air-conditioning systems. Int J Heating, Ventilating, Air-Cond Ref Res 5(4):353–380
3. Hung CP, Wang MH (2003) Fault diagnosis of air-conditioning system using CMAC neural network approach, advances in soft computing, engineering design and manufacturing. Springer, Heidelberg p 1
4. Wang S, Xiao F (2004) AHU sensor fault diagnosis using principal component analysis method, energy and buildings. Int J Refrig 36(2):147
5. Kima M, Yoonb SH, Domanskib PA, Vance W (2008) Design of a steady-state detector for fault detection and diagnosis of a residential air conditioner, PayneVolume 31. Issue 5:790
6. Steinwart I, Christmann A (2008) Support vector machine. Springer, Heidelberg
7. Cristianini N, Shawe Taylor J (2000) An introduction to Support vector machines and other kernel-based learning methods. Cambridge university press, Cambridge
8. Kawai T, Kushizaki S (2010) Diagnosis of air-conditioner By SVM. In: Proceedings of the 23rd International Congress on Condition Monitoring and Diagnostic Engineering Management, p 575
9. Michael MT (2001) Introduction to physical modeling with Modelica. Kluwer Academic Publishers, Netherland

Fault Detection and Remaining Useful Life Estimation Using Switching Kalman Filters

Reuben Lim and David Mba

Abstract The use of condition monitoring (CM) data in degradation modeling for fault detection and remaining useful life (RUL) estimation have been growing with increasing use of health and usage monitoring systems. Most degradation modeling methods requires fault detection thresholds to be established. When the CM measure exceeds the detection threshold, RUL prediction is then performed using a time-invariant dynamical model to represent the degradation path to the failure threshold. Such approaches have some limitations as detection thresholds can vary widely between individual units and a single dynamical model may not adequately describe a degradation path that evolves from slow to accelerated wear. As such, most degradation modeling studies only focuses on segments of their CM data that behaves close to the assumed dynamical model. In this paper, the use of Switching Kalman Filters (SKF) is explored for both fault detection and remaining useful life prediction under a single framework. The SKF uses multiple dynamical models describing different degradation processes from which the most probable model is inferred using Bayesian estimation. The most probable model is then used for accurate prediction of RUL. The proposed SKF approach is demonstrated to track different evolving degradation path using simulated data. It is also applied onto a gearbox bearing dataset from the AH64D helicopter to illustrate its application in a practice.

Keywords Switching kalman filter · Fault detection · Degradation modeling · Remaining useful life

R. Lim (✉)
Senior Engineer, Republic of Singapore Air Force, Singapore, Singapore
e-mail: r.limchikeong@cranfield.ac.uk

D. Mba
Head of Turbo-Machinery and Icing Group, Cranfield University, Bedford , UK
e-mail: d.mba@cranfield.ac.uk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_6

1 Introduction

The prevalence of health and usage monitoring systems (HUMS) on aircrafts in the past decades has fuelled the growth of using condition monitoring (CM) data in degradation modeling for health assessment of critical systems [1]. A wide variety of approaches in use of CM data for degradation modeling were comprehensively reviewed by different authors [1–4]. These data driven approaches are broadly classified into three main categories namely, physics-based methods, artificial intelligence methods and statistical methods. Physics-based method uses deterministic model of the system and can be very complex to develop. Artificial intelligence methods such as neural networks and support vector machines can handle highly non-linear problems but it requires huge number of training data which is often not available in practice. Amongst the three, statistical method is the most widely used in industry where conventional statistical process control and trend extrapolation are most commonly applied [3]. Advanced statistical methods such as Hidden Markov Models and cluster analysis can classify faults better but not widely used in practice due again to unavailability of training data. Most of the applications in the literature used experimental or simulated data for model training and little work was done with fielded applications [4]. In this paper, the switching Kalman filters (SKF) is investigated for fault detection and remaining useful life (RUL) of rolling element bearing and applied to both simulated and actual CM data gathered from AH64D helicopter.

2 Literature Review

The Kalman filter is a stochastic filtering process, which recursively estimates the state of a dynamic system in the presence of measurement noise and process noise, by minimizing the mean squared error [4]. The Kalman filter has been a widely applied concept in navigation and is also used in fields such as signal processing and econometrics. The Kalman Filter requires less training data compared to other statistical and AI techniques as it relies mainly on individual system's measurement data. However, the dynamical behavior of the system is required and be represented as a state-space model. In prognostic application, the Kalman filter was applied to predict the RUL of electrical connections [5] and electrolytic capacitors [6]. In these applications, the Kalman filter was used to adaptively track changes in the degradation process of the system and the dynamical model describing the degradation process was assumed to be time invariant. However, the degradation process in components can be uncertain and evolve over time as seen in bearing wear tests [7]. For example, in Fig. 1, the vibration measurement of a serviceable bearing can be stationary with measurement noise. When slow stable wear from damage such as surface pitting occurs, the vibration can gradually rise as a linear function. When accumulated damage is severe and unstable, the vibration rises rapidly in higher order functions.

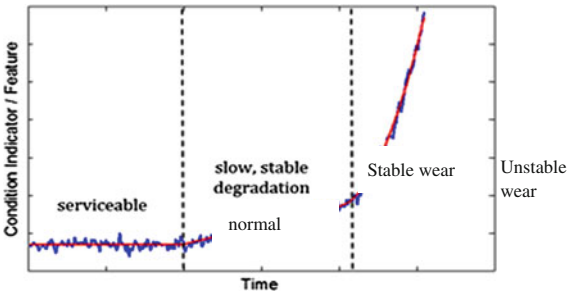


Fig. 1 Evolution of degradation process across time

As such, a single dynamical model may not adequately represent the different degradation processes. Consequently, this can cause predictions to diverge or fluctuate depending on whether the degradation process is under or over-fitted. This constraint is often seen in works [8, 9] where only measurements above an established threshold are considered in the analysis as those below does not behave according to the assumed dynamical model. For such problems, SKF can track the dynamics of the degradation process as it changes. RUL prediction is then performed based on the most probable dynamical model representing the degradation process. To do this, the SKF consists of multiple linear state-space models; like the basic Kalman filter, and it can switch between these models through a weighted combination across time. It is popularly used to track multiple moving targets but has also been applied in meteorology [10] and econometric [11]. SKF is applied here to track the different bearing degradation processes shown in Fig. 1. By tracking the dynamical behavior of different degradation processes, fault detection can be performed without using pre-established detection thresholds. It also helps maintainers to predict RUL more accurately by distinguishing between stable and unstable wear and performing prediction only when unstable wear is detected.

3 Background

This section provides a brief review of the Extended Kalman filter, dynamic Bayesian network, SKF and their application towards fault detection and RUL estimate of rolling element bearing.

3.1 Extended Kalman Filter

As mentioned, the Kalman filter recursively estimates the state mean and covariance of a linear process by minimizing the mean square error. The Extended Kalman filter (EKF) is a non-linear extension which uses linear approximation of the non-linear function to estimate the state mean and covariance [12, 13].

The linear approximation performed through first and second-order taylor series expansion of the non-linear function is most commonly used and the first-order is adopted here. The discrete state-space model describing a non-linear process is given by:

$$x_t = f(x_{t-1}) + q_{t-1},\ y_t = h(x_t) + r_t \tag{1}$$

where x_t is the true but hidden state of the system and y_k is the observable measurement of the state. *f(.)* is the fundamental matrix describing the system dynamics and *h(.)* is the measurement matrix and both are functions assumed to be continuously differentiable. $q_{t-1} \sim N(0, Q_t)$ is the process noise and $r_{t-1} \sim N(0, R_t)$ is the measurement noise. The EKF estimates the value of x_t, given the measurement, y_t by filtering out the noises. This is carried out using the 'Prediction' and 'Update' steps also known as the Ricatti Equations [13] are shown as follows.

Prediction Step:

$$\begin{aligned} \text{Predicted state estimate}: &\quad \hat{x}_t = f(x_{t-1}, t-1) \\ \text{Predicted estimate covariance} &\quad \hat{P}_t = F(x_{t-1}, t-1)P_{t-1}F^{'}(x_{t-1}, t-1) + Q_{t-1} \end{aligned} \tag{2}$$

Update Step:

$$\begin{aligned} \text{Measurement residual}: &\quad v_t = y_t - h(\hat{x}_{t-1}, t) \\ \text{Residual covariance} &\quad C_t = H(\hat{x}_t, t)\hat{P}_t H^{'}(\hat{x}_t, t) + R_t \\ \text{Kalman Gain} &\quad K_t = \hat{P}_t H^{'}(\hat{x}_t, t)C_t^{-1} \\ \text{Updated state estimate} &\quad x_t = \hat{x}_t + K_t v_t \\ \text{Updated estimate covariance} &\quad P_t = (I - K_t H(\hat{x}_t, t))\hat{P}_t \end{aligned} \tag{3}$$

where *F(.)* and *H(.)* are the Jacobians of *f*(.) and *h(.)* are given by

$$F(x_{t-1}, t-1) = \left.\frac{\partial f(x_{t-1}, t-1)}{\partial x}\right|_{\hat{x}_{t-1|t-1}},\ H(\hat{x}_t, t) = \left.\frac{\partial h(x_t, t)}{\partial x}\right|_{\hat{x}_{t|t-1}}, \tag{4}$$

3.2 Switching Kalman Filter

The switching Kalman filter may be represented as a dynamic Bayesian network. In each time step, both the model switch variable, S_t and state variable, x_t are hidden and have to be inferred from the observations, y_t. For a system with multiple dynamics which are described with n Kalman filters, the size of the belief state will increase exponentially at each time step to n^t. As such, inferring the probability of every state at each time step becomes intractable. To overcome this problem,

approximation method like the Generalised Pseudo Bayseian (GPB) algorithm as described in [12] was adopted. In each time step, the state and covariance estimates from all the filters in the previous time step are combined with weights assigned according to the mix probabilities of the model switch variable, $S_t^{i|j}$ and the model transition probability, Z_{ij} as shown in Eqs. (5) and (6).

$$\text{Model switching probabilities:} \quad S_t^{i|j} = \frac{Z_{ij} S_{t-1}^i}{\sum_{i=1}^n Z_{ij} S_{t-1}^i} \tag{5}$$

Weighted state and covariance estimates:

$$\tilde{x}_{t-1}^j = \sum_{i=1}^n S_t^{i|j} x_{t-1}^i, \; \tilde{P}_{t-1}^j = \sum_{i=1}^n S_t^{i|j} \left\{ P_{t-1}^i + \left[x_{t-1}^i - x_{t-1}^j \right] \left[x_{t-1}^i - x_{t-1}^j \right]' \right\} \tag{6}$$

with the weighted state and covariance estimates, the usual Kalman filter as shown in Eqs. (2) and (3) is carried out for each filter model with each yielding a predicted state, $\hat{x}_{t-1}^j$ and covariance, $\hat{P}_{t-1}^j$ estimate. The likelihood of each filter is then determined with Eq. (7) using their measurement residual, v_t^i. The probability of each model at the current time step can then be obtained as shown in Eq. (8). The weighted state and covariance estimate update for the current time can also be determined using Eq. (9). A detailed description of SKF is available in [14] and a good demonstration of SKF with use of GPB is shown in [15].

$$\text{Likelihood of filter from measurement residual:} \quad L_t^i = N(v_t^i; 0, C_t^i) \tag{7}$$

Probability of each model:

$$S_t^i = \frac{L_t^i \left(\sum_{i=1}^n Z_{ij} S_{t-1}^i \right)}{\sum_{i=1}^n \left(L_t^i \sum_{i=1}^n Z_{ij} S_{t-1}^i \right)} \tag{8}$$

The weighted state and covariance estimate update are computed as follows:

$$x_t = \sum_{i=1}^n S_t^i x_t^i, \; P_t = \sum_{i=1}^n S_t^i \left\{ P_t^i \left[x_t^i - x_k \right] \left[x_{t-1}^i - x_t \right]' \right\} \tag{9}$$

4 SKF Formulation for Tracking Varying Degradation Processes

In this analysis, it is assumed that component degradation is monotonically increasing and it evolves from normally operating to stable wear and then unstable wear. For bearings, a linear, polynomial or exponential model is used to describe

the different trends in the vibration-based degradation measure [16–18]. A Kalman filter is built for each of them and they are used together in the SKF. For the exponential filter, extended Kalman filter is applied due to its non-linear form. The state transition $F_i(.)$ is obtained from the Jacobian of the state equations using Eq. (4). It is assumed that the process noise entering the system only consists of zero mean white noise q_a and q_b which models the wear rate parameters a_t and b_t stochastically. The state, transition and process noise covariance for each filter are shown below with subscripts 1, 2 and 3 denoting the zero, first order and exponential Kalman filters respectively.

Zero Order polynomial model (Normal Operation)

$$\begin{aligned} \text{State} &: x_t = x_{t-1} \\ \text{State Transition} &: F_{1,t} = 1 \\ \text{Process Noise} &: Q_{1,t} = 0, y_t = x_t + r_t \\ \text{Measurement} &: H_{1,t} = 1 \end{aligned} \tag{10}$$

1st Order polynomial model (Stable Wear)

$$\begin{aligned} \text{State} &: x_t = x_{t-1} + a_{t-1}\Delta t,\ a_t = a_{t-1} + q_a \\ \text{State Transition} &: F_{2,t} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \\ \text{Process Noise} &: Q_{2,t} = \begin{bmatrix} 0 & 0 \\ 0 & q_a \end{bmatrix} \\ \text{Measurement} &: y_t = x_t + r_t,\ H_{2,t} = \begin{bmatrix} 1 & 0 \end{bmatrix}' \end{aligned} \tag{11}$$

Exponential model (Unstable Wear)

$$\begin{aligned} \text{State} &: x_t = x_{t-1} e^{b_{t-1}\Delta t},\ b_t = b_{t-1} + q_b \\ \text{State Transition} &: F_{3,t} = \begin{bmatrix} e^{b_{t-1}\Delta t} & x_{t-1}\Delta t e^{b_{t-1}\Delta t} \\ 0 & 1 \end{bmatrix} \\ \text{Process Noise} &: Q_{3,t} = \begin{bmatrix} 0 & 0 \\ 0 & q_b \end{bmatrix} \\ \text{Measurement} &: y_t = x_t + r_t,\ H_{3,t} = \begin{bmatrix} 1 & 0 \end{bmatrix}' \\ \text{Model transition matrix} &: Z = \begin{bmatrix} 0.99 & 0.005 & 0.005 \\ \sim 0 & 0.99 & 0.01 \\ \sim 0 & \sim 0 & \sim 1 \end{bmatrix} \end{aligned} \tag{12}$$

Initial model probabilities, state and covariance estimate:

$$S_0 = [0.98 \quad 0.01 \quad 0.01], x_0 = y_{0,} a_0 = 0, b_0 = 0, P_0 = I \tag{14}$$

For the SKF, the state transition matrix Z is set such that the system tends to remain in its own state with $Z_{ii} \sim 1$. It is also assumed that the degradation rate can only progress i.e. from normal to stable and unstable degradation but not the reverse. However, Z_{ij} is assigned a value approximately zero for $i > j$ as a value of zero can cause underflow problems in Eq. (8) when implemented as a software program. The initial model probability, S_0 is set with high probability that its in normal condition. The initial state estimate, x_0 is initialized to the first measurement and initial parameters a_0 and b_0 are zero. The initial covariance matrix, P_0 is set arbitrarily with an identity matrix, I.

5 Diagnostics of Evolving Degradation Processes Using Simulated Data

The SKF approach to track the degradation processes is demonstrated here using simulated data. Figure 2 shows different evolving degradation processes; (1) normally operating to unstable wear at t = 150 h and (2) normally operating to stable wear at t = 100 h and then unstable wear at t = 200 h. The simulated degradation measurements are generated using the measurement equations from Eqs. (11–13). An additive measurement noise, $r \sim N(0, 0.08^2)$ is added all three processes. For stable wear, a wear rate parameter, a = 0.01 is adopted with process noise,

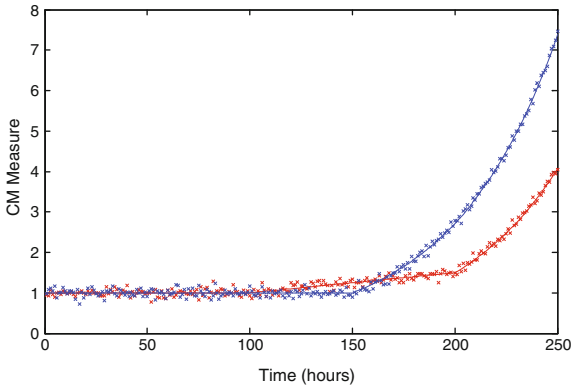


Fig. 2 Simulated degradation processes with measurement and process noise: (*1*) normally operating to unstable wear at t = 150 h and (*2*) normally operating to stable wear at t = 100 h and unstable wear at t = 200 h

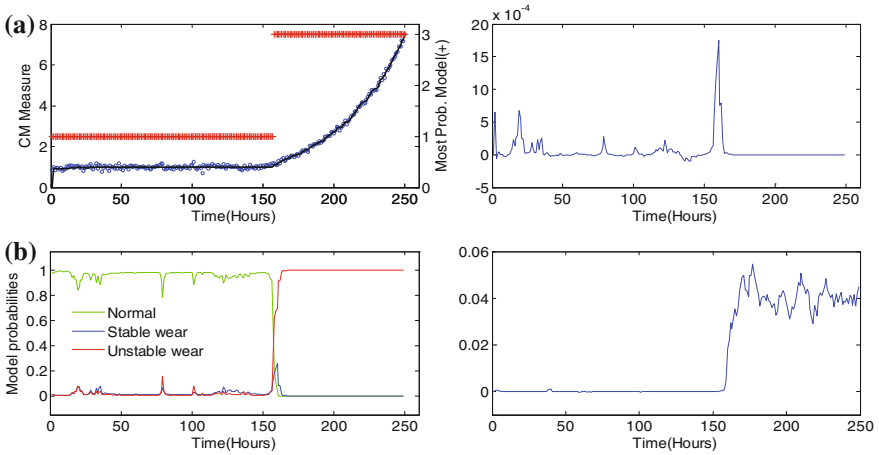


Fig. 3 Normal to unstable wear (*Top left*) Filtered state and most probable model (*Bottom left*) Model probabilities, (*Top* and *bottom right*) Estimated parameters a_t and b_t

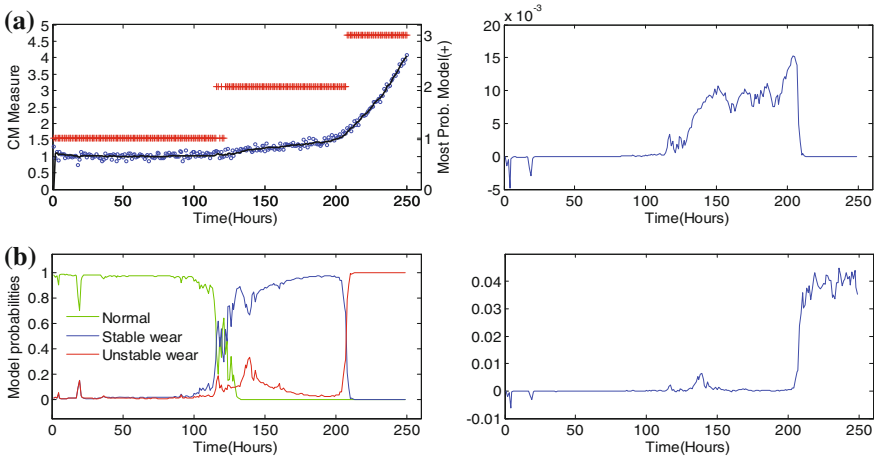


Fig. 4 Normal to stable and unstable wear (*Top left*) Filtered state and most probable model, (*Bottom left*) Model probabilities, (*Top* and *bottom right*) Estimated parameters a_t and b_t

$q_a \sim N(0, 0.001^2)$. For unstable wear, a wear rate parameter, $b = 0.04$ is adopted with process noise, $q_b \sim N(0, 0.004^2)$.

The ideal case where the dynamical models of the degradation processes and their measurement and process noise are known is shown here. Figures 3 and 4 shows the SKF results in tracking the evolving degradation processes. It can be seen that the SKF is able to track and estimate the most probable degradation process well using the dynamical behavior of the measurement. For normal to unsteady

wear, the SKF detects the change at 158 h compared to 150 h. For normal to steady and then unsteady wear, the SKF detects the change at 116 h and 208 h compared to 100 h and 200 h respectively. The SKF lags behind the actual transition times as it is performing the estimation in real-time and requires adequate measurements from the dynamical process. In addition, it can estimate the wear rate parameters, a and b well at ~ 0.001 and ~ 0.04. It should be noted that the estimation will not converge towards the exact parameter value due to inherent noise added to the measurements.

6 Case Study on AH64D Helicopter Tail Rotor Gearbox Bearing

The SKF approach is applied to vibration CM data from the AH64D Tail Rotor Gearboxes (TRGB) in a practical scenario. The bearing CM data and results from the SKF is shown in Fig. 5. The measurement error, r = 3.2e−4 is obtained by taking the variance of the stationary measurements when the TRGB is in a good condition and this can vary between individual gearboxes. The process error, q_s contains the uncertainty of the filters in modeling the real world [13]. It is obtained by tuning the SKF model with similar defect cases and is assumed to be the same across gearbox bearings. The SKF formulations are applied with q_s set initially as a small percentage of the measurement error, R. The SKF model is then applied on the CM data and q_s is tuned till the model is acceptably consistent yet responsive to changes in the degradation processes. In this study, q_s = *5e-8* is obtained by tuning the model using CM data from other TRGB with similar failure. Form Fig. 5, it can

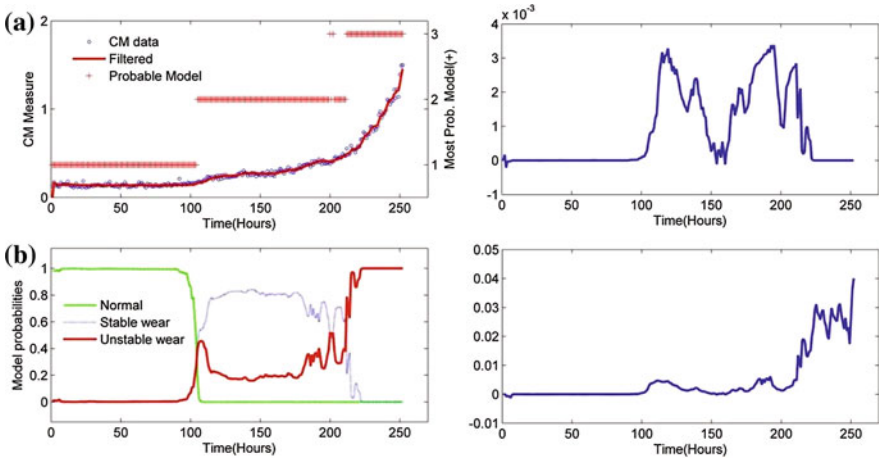


Fig. 5 TRGB CM data (*Top left*) filtered state and most probable model, (*Bottom left*) Model probabilities, (*Top* and *bottom right*) Estimated parameters a_t and b_t

be seen that the SKF can adaptively track the different bearing degradation processes with the process noise tuned from other gearboxes. However, when the CM measurements are not increasing monotonically at ~200 h, the SKF has to take a longer time before it converges. Instead of relying on the absolute value of the CM measurements, the SKF uses the dynamic behavior between the current and past measurement to diagnose the degradation state. Therefore, it is not dependent on a fixed threshold which are typically derived from statistical evaluation of large numbers of past failure cases. Another key advantage of this technique for diagnosis is that it provides the probability of which degradation process the bearing is in. In comparison, the widely used, statistical process control (SPC) approach only triggers when the measurement is above a statistical limit and no further information is available. The quantitative probability measure from the SKF allows more support for maintenance engineers as the probabilities of the bearing conditions can be compared in the event of an outlier measurement.

7 Prediction of Remaining Useful Life

The SKF infers the most probable dynamic model to be applied at each time step for prediction and the RUL of the bearing is predicted whenever an unsteady wear is detected. The RUL is predicted by propagating the weighted state and covariance estimates obtained from Eq. (8) at each time step using Eq. (2) and determining the time when the degradation state crosses the failure threshold. The α–λ metric [19] is applied to evaluate the performance of this prognostic evaluation as shown in Fig. 6.

The α-λ metric compares the actual RUL to the predicted RUL with converging α bounds that provides an accuracy region. The α bounds are application specific and a prediction is correct if it falls within the alpha bounds. From Fig. 6, the prognostic algorithm performs well as its accuracy improves quickly with time within the 30 % bounds. However, there are points on the RUL trajectory that lies outside the accuracy zone towards the end of useful life which is a behavior

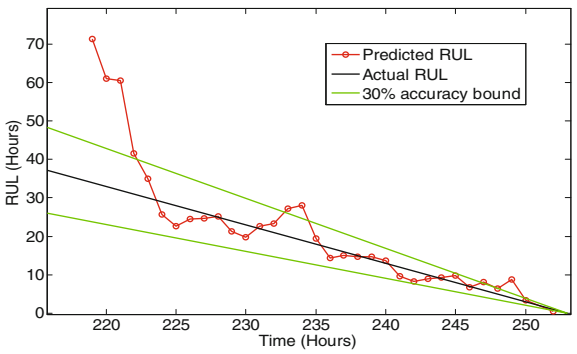


Fig. 6 α–λ performance metric using 30 % accuracy bounds

reportedly observed in [20] as well. This behavior could be attributed to unsteady vibration levels as the accumulated damage in the bearing becomes sizeable and could perhaps be addressed by lowering the failure threshold limit. Besides the RUL estimate, most of the lower confidence bound, which is important for conservative estimate of the RUL prediction are close to the lower 30 % accuracy bound as well.

8 Conclusion

In this study, the use of SKF is applied for fault detection and RUL estimation. The method is applied to both simulated data and actual helicopter gearbox bearing with promising results. The SKF model allows for degradation processes to evolve through time from which the underlying dynamical process would be inferred accordingly. The advantages of this approach are that it does not depend on a fixed threshold for fault detection and it can model the different degradation processes as they evolve. This approach also provides maintainers with more information for decision-making as a probabilistic measure of the state of bearing degradation is available. From the prognostic performance metric, it was shown that the RUL estimates have high accuracy when it is inferred that the degradation process is likely to be unstable. This in turn can provide maintainers with higher confidence on the predicted RUL for maintenance planning. A drawback of this method is that it requires frequent acquisition of measurement for the filter estimation which may not be readily available in practice.

References

1. Gorjian N, Ma L, Mittinty M, Yarlagadda P, Sun Y (2009) A review on degradation models in reliability analysis. In: Proceedings of the 4th World Congress on Engineering Asset Management. Springer, Athens
2. Heng A, Zhang S, Tan ACC, Mathew J (2009) Rotating machinery prognostics: State of the art, challenges and opportunities. Mech Syst Signal Process 23(3):724–739
3. Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 20(7):1483–1510
4. Sikorska JZ, Hodkiewicz M, Ma L (2011) Prognostic modelling options for remaining useful life estimation by industry. Mech Systems Signal Process 25(5):1803–1836
5. Lall P, Lowe R, Goebel K (2010) Prognostics using Kalman-Filter models and metrics for risk assessment in BGAs under shock and vibration loads. In: Proceedings 60th Electronic Components and Technology Conference (ECTC), 2010 pp. 889
6. Jose RC, Chetan K, Gautam B, Sankalita S, Kai G (2011) A Model-based Prognostics Methodology for Electrolytic Capacitors Based on Electrical Overstress Accelerated Aging. In: Annual Conference of the Prognostics and Health Management Society
7. Kotzalas MN, Harris TA (2001) Fatigue failure progression in ball bearings. J Tribol 123 (2):238–242

8. Gebraeel N (2006) Sensory-updated residual life distributions for components with exponential degradation patterns. Autom Sci Eng IEEE Trans 3(4):382–393
9. Roulias D, Loutas TH, Kostopoulos V (2012)A hybrid prognostic model for multistep ahead prediction of machine condition. J Phys: Conf Series. 364(1):012081
10. Manfredi V, Mahadevan S, Kurose J (2005) Switching kalman filters for prediction and tracking in an adaptive meteorological sensing network. In: Sensor and Ad Hoc Communications and Networks, 2005. IEEE SECON 2005. 2005 Second Annual IEEE Communications Society Conference on, pp 197
11. Lim Y, Cheng S (2012) Knowledge-driven autonomous commodity trading advisor. 2012 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, Macau
12. Särkkä S (2011) Optimal Filtering. In: Bayesian Estimation of Time-Varying Systems: Discrete-Time Systems pp 36–38
13. Zarchan P, Musoff H (2005) Polynomial Kalman filters. In: Fundamentals of Kalman Filtering: A Practical Approach, 2nd ed, American Institute of Aeronautics and Astronautics, USA, pp 156
14. Kevin M (1998) Learning switching Kalman filter models, 98-10. Compaq Cambridge Research Lab Tech Report
15. John Q SKF Demo. http://www.cit.mak.ac.ug/staff/jquinn/software.html (accessed October/21)
16. Gebraeel N, Lawley M, Liu R, Parmeshwaran V (2004) Residual life predictions from vibration-based degradation signals: a neural network approach. Indus Elect IEEE Trans 51(3):694–700
17. Shao Y, Nezu K (2000) Prognosis of remaining bearing life using neural networks. In: Proceedings of the Institution of Mechanical Engineers, Part I. J Syst Control Eng 214(3):217
18. Sutrisno E, Oh H, Vasan A. SS, Pecht M (2012) Estimation of remaining useful life of ball bearings using data driven methodologies. In: Prognostics and Health Management (PHM), 2012 IEEE Conference on, pp 1
19. Saxena A, Celaya J, Balaban E, Goebel K, Saha B, Saha S, Schwabacher M (2008) Metrics for evaluating performance of prognostic techniques. In: Prognostics and Health Management, 2008. PHM 2008. International Conference on, pp 1
20. Saxena A, Celaya J, Saha B, Saha S, Goebel K (2009) On applying the prognostic performance metrics. In: International Conference on Prognostics and Health Management (PHM)

A Novel Integrated Sensor for Stress Measurement in Steel Strand Based on Elastomagnetic and Magnetostrictive Effect

Xiucheng Liu, Bin Wu and Cunfu He

Abstract The deteriorating state of steel strands can be assessed using information of its overall and local stress level. An integrated sensor, which is based on magnetostrictive and elastomagnetic effect, is proposed for stress measurement of seven-wire steel strands. This integrated sensor can alternatively measure the stress level of seven-wire steel strands through guided wave-based technique and elastomagnetic method. When it works as magnetostirctive sensor to generate longitudinal guided wave in strand, the missing frequency band occurred in received wave signal is investigated in a pre-tensioned strand. It is found that the missing band central frequency increases linearly as the value of natural logarithm of overall stress increases. The integrated sensor can be applied to measure local stress in a strand based on elastomagnetic effect. The output signal amplitude of the sensor has linear relationship with strand stress level and stress increment of less than 10 MPa can be identified. The characteristic functions for both local and overall stress measurement are pre-calibrated experimentally. The proposed integrated sensor has great potential for applications of steel strands stress measurement.

1 Introduction

High tensile steel strands are commonly used in prestressed structures such as suspension bridges and concrete buildings. Maintenance of such infrastructures becomes one of the major concerns in civil engineering area. The ability to monitor the integrity of steel strands can help decide when to replace them and how to prolong their useful lifetime.

During last decades, many efforts have been made to develop reliable non-destructive testing (NDT) methods for stress measurement and defect detection in strand/cable structures. Magnetostrictive-based guided wave technology has been

X. Liu (✉) · B. Wu · C. He
Beijing University of Technology, Pileyuan 100, Beijing 100124, China
e-mail: xiuchliu@bjut.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_7

applied for flaws detection in long steel strands and cables [1, 2]. Based on magnetostrictive-based transduction system, experimental tests were conducted to investigate the effect of tensile load on the guided wave signal propagating in a strand. The presence of a missing frequency band in the received signals' spectrums indicated that the guided wave energy attenuation was related to the applied tensile force. The central frequency of the missing band was observed to regularly shift as the increasing tensile force in the strand [3]. Several researchers focused on revealing the shifting characteristics of such missing frequency band with theoretical research methods [4, 5]. From another perspective, if the missing frequency band can be determined with experimental techniques, then a new stress measurement method can be developed. The direct transmitter-to-receiver wave signal was selected for extracting the missing frequency band information, so that the guided wave-based method can measure the overall stress level of the structures [6].

Both local stress and overall stress level can be used for the assessment of strands' deteriorating state. Among the available methods that are capable of measuring local stress level of strands/cables, elastomagnetic sensor (EMS) based technique is the most promising one [7, 8]. The EMS was designed based on the principle that the permeability of ferromagnetic materials is a function of applied fields (magnetic field, stress and temperature), and this characteristic function can be experimentally pre-calibrated [9].

Although the guided wave- and EMS-based technique can individually measure the overall or local stress, each of them cannot meet the requirements of comprehensively evaluating the strand/cable deteriorating state. This paper will present a novel integrated sensor structure to make it possible that a single sensor can alternatively measure the overall and local stress of seven-wire steel strands.

The rest of paper is organized as follows. Measuring principle and the integrated sensor configuration is given in Sect. 2. Experimental results are presented for measuring overall and local stress in Sects. 3 and 4, separately. Finally, our conclusions are given in Sect. 5.

2 Measuring Principle and Integrated Sensor

Magnetostictive effect and its inverse effect (or elastomagnetic effect) refer to the coupling phenomenon between magnetic field and mechanical deformation in ferromagnetic materials. Various types of sensors are developed based on these two physical effects, such as magnetostrictive sensor (MsS) for ultrasonic guided wave excitation and EMS for permeability measurement in cylindrical waveguides [10, 11]. Based on the structures of conventional MsS and EMS, here an integrated sensor configuration is devised for steel strand stress measurement. The structure of the integrated sensor is shown in Fig. 1a which includes a bias magnetic circuit, an outer sensing coil and an inner sensing coil. The bias magnetic circuit is made of permanent magnets and yokes to provide static magnetic field for material magnetization. The axial length of the outer sensing coil is more than three times of that of the inner

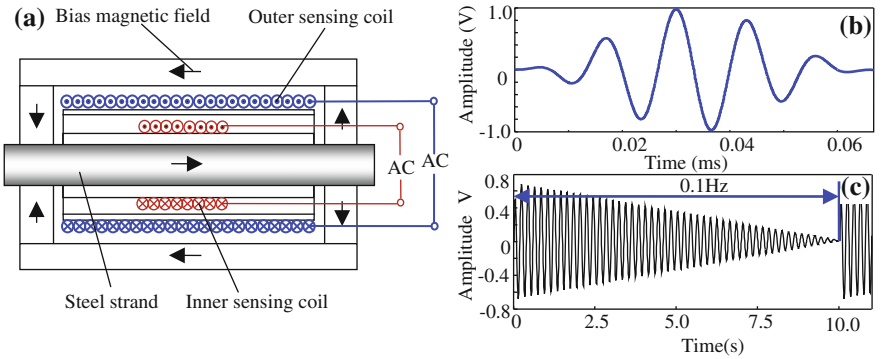


Fig. 1 a the integrated sensor configuration diagram and its typical excitation signal waveform for b MsS mode and c EMS mode

sensing coil. By having such configurations, the integrated sensor can work as MsS or EMS.

When the integrated sensor works as a MsS, both the outer and inner sensing coil can be selected as transmitter coil. After feeding sinusoidal tone burst signal modulated by a Hanning window into the sensing coil, ultrasonic longitudinal guided waves can be generated in steel strands. When the guided wave propagates along the strand for a certain distance, another receiver sensor will receive the direct transmitter-to-receiver wave (DTRW) signal. Thus, the missing frequency band information can be extracted from the DTRW signal for the characteristic function determination. Consequently, the overall stress level of the strands can be derived using measured missing band information and the characteristic function.

The EMS has several implementation modes, in this paper a sinusoidal current with a linearly decaying envelope, which is shown in Fig. 1c, was applied to the outer sensing coil [12]. A digital oscilloscope acquires the output signals of inner sensing coil while the tension force is different. For a given EMS, its output signal amplitude will be a function of the applied stress according to the elastomagnetic effect. Similarly, the local stress can be calculated by substituting the output signal amplitude into the pre-calibrated formula.

As mentioned previously, the formula for calculating both the overall and local stress must be pre-calibrated. A steel strand stretching system, which is shown in Fig. 2, is employed for the calibration experiments. Two integrated sensors are mounted onto the tested strand. A hydraulic cylinder is configured to provide anaxial load to the strand from zero tons to about twenty tons.

It is noted that the axial load applied to the strand is assumed to be evenly distributed over the whole bundle of steel wires. Thus, the averaged overall stress can be calculated with the axial force and the net cross-sectional area of the seven wires. The actual axial force of the strand is measured by a calibrated load cell. All the following experimental results are obtained by using this steel strand stretching system and the proposed integrated sensor.

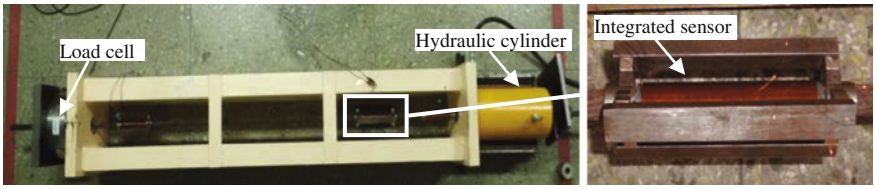


Fig. 2 The picture of steel strand stretching system (Left) and the integrated sensor (Right)

3 Overall Stress Measurement Based on Magnetostrictive Effect

The distance of the two integrated sensors as shown in Fig. 2 is fixed at about 0.6 m. The central and helical wire of the tested strand has a diameter of about 6.3 and 6.0 mm separately. A 5-cycle sinusoidal tone burst modulated by a Hanning window as shown in Fig. 1b is fed into the inner transmitter coil. The central frequency of the excitation tone burst is about 75 kHz. Longitudinal guided wave can be generated in the strand based on magnetostrictive process. Another sensor can receive the DTRW signals, the typical received signal waveforms are shown in Fig. 3a.

As indicated in Fig. 3a, when the tensile stress, S, of the strand increases from 0 to 488 MPa, the degree of dispersion of the guided wave gets aggravated and this results in the presence of several overlapped wave packets in DTRW signal. The short time Fourier transform (STFT) technique can extract the dispersion characteristics in a certain frequency range, and the STFT results of time-domain signals

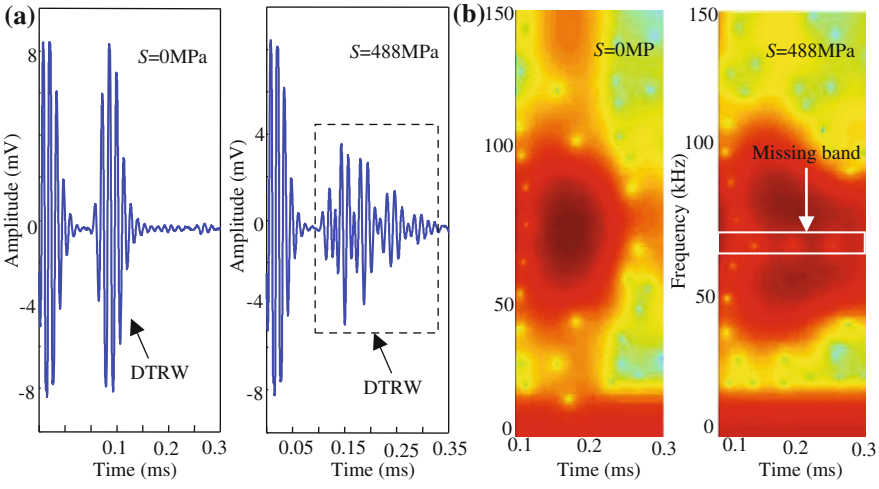


Fig. 3 a the received signals in time domain and b its STFT results

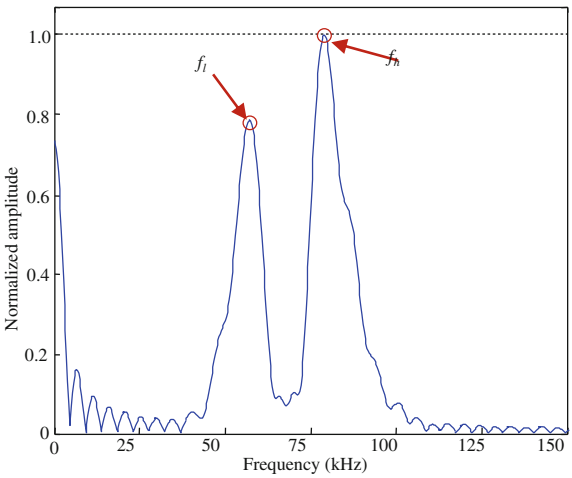


Fig. 4 Fourier transforms result of the received DTRW signal when the tensile stress is about 488 MPa

are shown in Fig. 3b. It is obviously that the propagation time interval for some frequency components increases. That is, the group velocities of guided wave at such frequencies become slower than that in a free strand case. The higher degree dispersion of guided wave is attributed to the intensified mechanical contact among the wires in a tensioned strand.

The distribution of guided wave energy in the excitation frequency range rearranges to have a narrow frequency band around 70 kHz in which the amount of energy experiences seriously attenuation. Figure 4 shows the Fourier transform (FFT) result of the received time domain signal while the tensile stress is S = 488 MPa. A missing frequency band can be observed at the spectrum of the DTRW signal. As mentioned previously, many researchers tend to believe that the characteristics of such missing frequency band are related to the complex geometry of strand and the complicated contact behaviors among the wires.

The two local extreme points aside the missing frequency band are concerned and their corresponding frequencies are marked as f_l and f_h. Then the central frequency of this missing band, f_c, can be simply calculated as follows:

$$f_c = (f_l + f_h)/2 \tag{1}$$

As reported in Ref. [4], this missing band shifts as the increase of tensile stress. The central frequency values for all tested cases are determined and plotted as a function of the applied tensile stress. As shown in Fig. 5, the value of f_c increases linearly with natural logarithm of stress. Their fitting equation is,

$$f_c = 8.5906 \times \ln(S) + 15.0089 \tag{2}$$

Finally, the function for overall stress measurement using MsS-based guided wave technique is characterized by Eq. (2). In practical applications, first apply

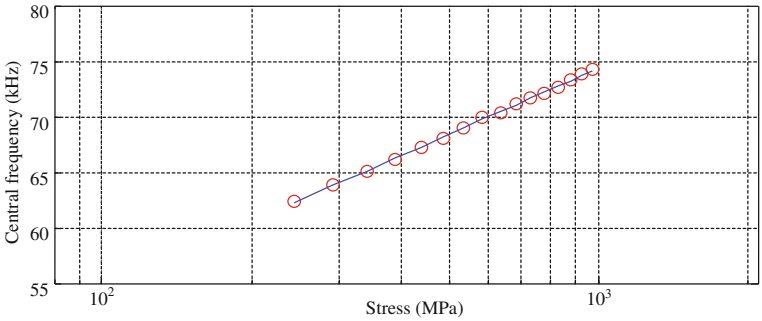


Fig. 5 The relationship between the central frequency of missing band and tensile stress of the strand

Fourier transform to the detected DTRW signal to find out the value of f_l and f_h. Second, the value of f_c can be calculated using Eq. (1), and then the tensile stress of strand can be determined by substituting the value of f_c into Eq. (2).

Although the proposed missing frequency band measurement-based technique can be applied for indirectly measuring the stress level in a strand, this method has some limitations in practice. For instance, in low stress level state the validity of Eq. (2) needs to be further investigated because the central frequency f_c is hard to determine from the received signal. In addition, the missing frequency band is decided by not only the tensile stress but also the strand structure and sizes. Therefore, theoretical research is an important prerequisite for estimating the central frequency f_c.

4 Local Stress Measurement Based on Elastomagnetic Effect

Each of the integrated sensors used for DTRW detection method can be applied for local stress measurement if it works as a MES. This time a sinusoidal current with a linearly decaying envelope is fed into the outer sensing coil. Thus, the strand material will experience magnetic disturbances. According to Lenz's law, the magnetic disturbances will induce voltage in inner sensing coil, and the acquired signal waveform is shown in Fig. 6a.

The peak-to-peak amplitudes of each cycle in the sinusoidal signal are extracted to form one of the ridge curves in Fig. 6b. This ridge curve is different from the linearly decaying envelope of excitation signal, but has nonlinear tendency in time domain. When tensile stress of the strands increases, the ridge curve will shift to have higher amplitudes. A fitting curve representing the relationship between the amplitude of ridge curve, U, and the stress can be obtained at each cycle of the excitation signal. However, the fitting curve that can be expressed as linear equation

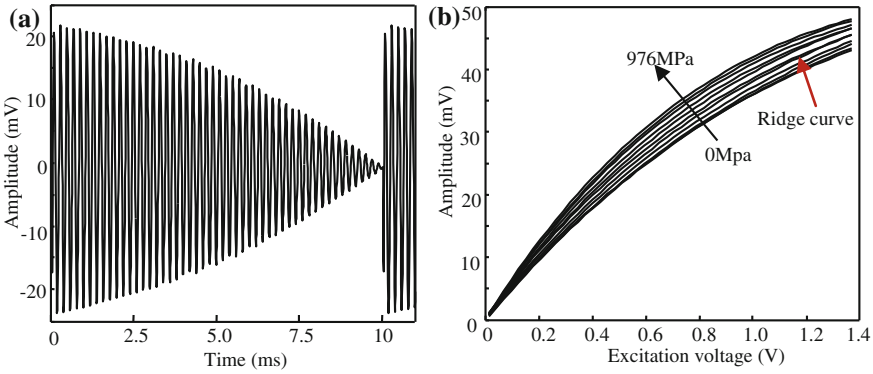


Fig. 6 a the waveform of typical signal received by the MES and b the extracted ridge curves for different tensile stress cases

is preferred for practical applications. Two criteria, stress measurement sensitivity, G, and linear regression coefficients, R, of the fitting curve, can be used for selecting the optimal excitation voltage.

The signal cycle with peak amplitude of 0.8163 V is selected for analysing the relationship between the stresses and the ridge curve amplitudes. The data of ridge curve amplitudes for different stress cases are plotted as open circles in Fig. 7. After applying least squares fit algorithm to these data, a linear regression equation is found to have following expression,

$$U = 8.0310S + 30.8458 \tag{3}$$

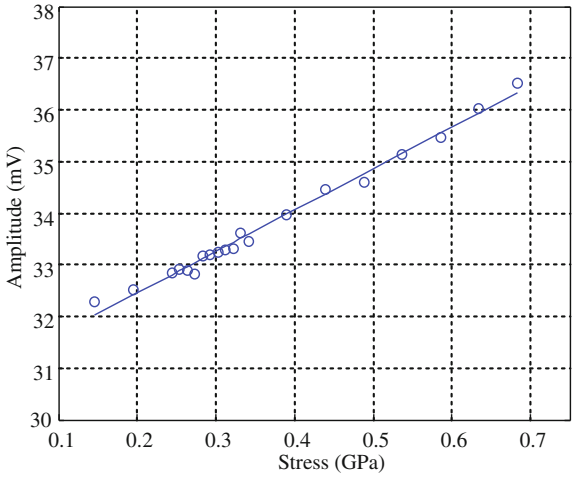


Fig. 7 The ridge curve amplitudes while tensile stress of the strand is different

The linear regression coefficients of the regression equation is about $R = 0.9936$. That means the value of U has a strong linear relationship with the increasing stress (in GPa). The stress measurement sensitivity of the sensor is about 8.0310 mV per GPa. Similarly, Eq. (3) can be treated as the characteristic function for local stress measurement using MES-based technique. The local stress value can be calculated by substituting the value of U into the Eq. (3).

In the range from 244 to 341.6 MPa, the stress increment is about 9.76 MPa. The experimental results obtained in this stress range well agree with the trend of the curve as illustrated in Fig. 7. That is, the minimum stress increment that can be identified by the proposed integrated sensor is less than 10 MPa.

5 Conclusion

A novel integrated sensor for stress measurement is proposed to measure both overall and local stress of steel strands. The implementation procedures for this integrated sensor are presented. The experimental results show that one single sensor can work as MsS and EMS alternatively to measure the overall and local stress of steel strands.

The characteristics of the missing frequency band occurred in DTRW signal is affected by the applied stress. The central frequency of the missing frequency band moves from the left to right in the spectrum of DTRW signal. When the stress reaches to a certain value, the missing frequency band will be out of the covered frequency range of the sensor. A broadband sensor would be helpful for observing the frequency band missing phenomena in higher stress cases. Although the central frequency f_c linearly increase with the increase of $\ln(S)$ and a new overall stress measurement method is illustrated in this paper, the precision and accuracy have to be further verified.

When the integrated sensor works as EMS, its stress measurement resolution is less than 10 MPa. With the given characteristic function for amplitude of ridge curve, the EMS-based techniques can be applied for local stress measurement with high accuracy.

Acknowledgments This work described in this paper was fully supported by a grant from the National Natural Science Foundation of China (Project No. 11132002).

References

1. Xu J, Wu XJ, Wang LY (2007) Detecting the flaws in prestressing strands using guided waves based on the magnetostrictive effect. Insight 49(11):647–650
2. Tse PW, Liu XC, Liu ZH et al (2011) An innovative design for using flexible printed coils for magnetostrictive-based longitudinal guided wave sensors in steel strand inspection. Smart Mate Struc 20(5):055001

3. Kwun H, Bartels KA, Hanley JJ (1998) Effects of tensile loading on the properties of elastic-wave propagation in a strand. J Acoust Soc Am 103(6):3370–3375
4. Treyssède F, Frikha A, Cartraud P (2013) Mechanical modeling of helical structures accounting for translational invariance Part 2: guided wave propagation under axial loads. Int J Solids Struct 50(9):1383–1393
5. Bartoli I, Castellazzi G, Marzani A et al (2012) Prediction of stress waves propagation in progressively loaded seven wire strands. In: Proceedings of SPIE-the international society for optical engineering, vol 8345. Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems
6. Di-Scalea FL, Rizzo P, Seible F (2003) Stress measurement and defect detection in steel strands by guided stress waves. J Mater Civ Eng 15(3):219–227
7. Ricken W, Schoenekess HC, Becker WJ (2006) Improved multi-sensor for force measurement of pre-stressed steel cables by means of the eddy current technique. Sens Actuators A Phys 129(1–2):80–85
8. Sumitro S, Kurokawa S, Shimano K et al (2005) Monitoring based maintenance utilizing actual stress sensory technology. Smart Mate Struc 14(3):68–78
9. Zhao Y, Wang ML (2006) Non-destructive condition evaluation of stress in steel cable using magnetoelastic technology. In: Proceedings of SPIE-the international society for optical engineering, vol 6178. Nonintrusive Inspection, Structures Monitoring, and Smart Systems for Homeland Security
10. Kwun H, Kim YS, Light GM (2003) The magnetostrictive sensor technology for long range guided wave testing and monitoring of structures. Mater Eval 61(1):80–84
11. Chen WM, Liu L, Zhang P et al (2010) Non-destructive measurement of the steel cable stress based on magneto-mechanical effect. In: Proceedings of SPIE-the international society for optical engineering, vol 7650(Part 1), Health Monitoring of Structural and Biological Systems
12. Lloyd GM, Singh V, Wang ML et al (2003) Temperature compensation and scalability of hysteretic/anhysteretic magnetic-property sensors. IEEE Sens J 3(6):708–716

Cavitation Sensitivity Parameter Analysis for Centrifugal Pumps Based on Spectral Methods

Kristoffer K. McKee, Gareth Forbes, Ilyas Mazhar, Rodney Entwistle, Melinda Hodkiewicz and Ian Howard

Abstract Cavitation is a major problem facing centrifugal pumps in industry today. Unable to constantly maintain operating conditions around the best efficiency point, centrifugal pumps are subject to conditions that may lead to vaporisation or flashing in the pipes upstream of the pump. The implosion of these vapour bubbles in the impeller or volute causes damaging effects to the pump. A new method of cavitation detection is proposed in this paper based on spectral methods. Data used to determine parameters were obtained under ideal conditions, while the method was tested using industry acquired data. Results were compared to knowledge known about the state of the pump, and the classification of the pump according to ISO 10816.

1 Introduction

Centrifugal pumps are found in many industries, pumping a range of fluids from fresh water to a slurry mixture. Due to their wide industrial applications, a large amount of research has been performed to attempt to identify fault modes, starting from their inception to complete fault development. While some faults are easy to detect, such as a leaking seal, others, such as cavitation, are not easily detectable until a later stage. Those faults that are not as easily detectable may render the machine close to inoperable when first detected.

This paper will focus on the detection of cavitation in a centrifugal pump, utilising results of octave band analysis on an acceleration signal as inputs to a neural network, to determine the state of the centrifugal pump as being one of three stages: no cavitation, incipient cavitation, or full-onset of cavitation. Cavitation, which is the formation of and subsequent implosion of vapour bubbles in a fluid,

K.K. McKee (✉) · G. Forbes · I. Mazhar · R. Entwistle · I. Howard
Department of Mechanical Engineering, Curtin University, Perth, Australia
e-mail: k.mckee@curtin.edu.au

M. Hodkiewicz
School of Mechanical and Chemical Engineering, The University of Western Australia, Perth, Australia

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_8

creates damaging effects on the centrifugal pump as the vapour bubbles implode on the vanes of the pump's impeller. Four main symptoms accompany the onset of cavitation: erosion via pitting of the impeller, a sharp crackling noise which is usually compared to the sound of "pumping rocks", high frequency and high amplitude vibration, and a reduction in pumping efficiency, [9, 11].

The following sections of the paper will discuss the state of the art in cavitation detection utilising neural networks, explain the theory behind the proposed methodology and procedure followed by the results obtained.

2 State of the Art in Cavitation Detection Using Neural Networks

Limited research has been done in the field of using neural networks to detect cavitation. An overall review of cavitation detection methods can be found in McKee et al. (McKee et al. 2011; [8]).

Wang et al. utilised a combination of wavelet analysis, rough sets, and partially linearized neural networks (PNN) to attempt to detect cavitation. The system was able to correctly detect cavitation with 85.1 % accuracy, [16, 18].

Wang et al. proposed a method using a B-spline neural network to detect fault modes, including cavitation, in a centrifugal pump. However, the method was only applied to seal ring faults with a 80 % success rate and not applied to cavitation, [3].

Nasiri et al. experimented with using a feed forward neural network to detect cavitation. 13 inputs collected from 3 accelerometer sensors (which were located in the front, radial and back of the volute) were given to the neural network: RPM of the pump, crest factor, kurtosis, mean and the variance of the signal. The system was shown to have a maximum error of 9.375 % when one sensor was used, a maximum error of 7.1875 % when 2 sensors were used, and 0 % error when all three sensors were used. Although providing very promising results, the number of sensors and inputs to the neural network are costly both in money for sensors and processing time, [2].

Other uses of neural networks to detect cavitation have resulted in mixed results. Most experiments have not been validated against measurements taken in industry, but instead data created in a lab under ideal conditions or created under simulations, [5, 19, 22]. As a result, its ability to perform on industrial equipment has not been verified.

3 Sensitivity Analysis Approach

Limited research has been done in the field of using neural networks to detect cavitation. An overall review of cavitation detection methods can be found in McKee et al. [8].

3.1 Octave Band Analysis

Octave band analysis divides the frequency spectrum of a signal into bands, or sections, which are then analysed separately to determine characteristics that may be otherwise hidden when the entire vibration signal is processed. Octave band analysis has been utilised in the acoustics field for decades for isolating and analysing specific regions of frequencies and has resulted in the creation of standards such as ISO 532. According to ISO 532, the upper and lower limits, and the central frequencies of the octave bands are predefined values, having the range cover the audible range of frequencies for humans (20 Hz–20 kHz). Upper and lower limits of each band are calculated to be the central frequency of the band divided by the square root of 2, and the central frequency of the band multiplied by the square root of 2. As a result, starting with a central frequency of 31.5 Hz, the entire audible range is covered in 10 octave bands, [13].

McKee et al. proposed an altered method in order to apply Octave analysis to the condition monitoring field, [7]. Since an important factor in rotating machinery vibration is the running speed of the rotating machine, and not the predefined range of frequencies analysed, two major changes were introduced to the method. First, the central frequency of the second octave band is set equal to the running speed of the rotating machinery. This causes the first octave band to contain only sub-harmonic information, and octave bands higher than the second to contain information surrounding the harmonics of the running speed. The second change is that the number of octave bands is not limited to 10 octave bands, but rather changes in order to accommodate half the sampling rate of the digitised vibration signal. As a result, the number of octave bands is dependent on the sampling rate and running speed of the machine. However, this also means that specific octave bands consistently contain information related to certain harmonics of the rotating machinery, which allows easy comparison of information of the same type of machines that maybe running at different speeds. For example, octave band 2 will contain information surrounding the running speed, while octave band 3 and 4 will contain information surrounding the 2nd and 3rd through 5th harmonic respectively, [7].

3.2 Principal Component Analysis

Principal Component Analysis (PCA) attempts to find a set of orthogonal axes that would maximize the amount of variance of the data. These orthogonal axes are a linear combination of the input variables, and are ordered according to the amount of spread of the variance along the axes. Hence, the first principal component is the axis along which the data contains the largest variance. The second principal component is the orthogonal axis that has the second largest variance, [12].

The goals of PCA are to [1, 21]

1. extract the most important information from the input data
2. remove the non-important data, resulting in a compressed data set
3. reduce or simplify the data set
4. determine if there are any relationships between objects.

As a result, the principal components obtained are arranged in the order of most to least importance in separating the data, if clusters exist within the data. This is the reasoning behind using this method at the pre-processing stage to a neural network.

3.3 Neural Network Approach

Artificial neural networks (ANN) are used widely for pattern recognition, since they perform well in separating clusters of data in noisy and unpredictable situations. A neural network contains a set of mathematical processing elements, also known as nodes, connected using synapses. Each node uses a weighted sum of the input information, which is obtained via connecting synapses, as the input to an activation function, which is usually either sigmoidal, arc tan or linear. The output of the activation function is then sent to connecting nodes to be used as the input of the subsequent node. If used individually, the application of a node is limited, however, when multiple nodes are inter-connected in layers, solutions to linear and nonlinear mathematical problems can be obtained with some training, [4, 10].

Training is performed by showing the network examples of input patterns along with the desired output patterns, and allowing the network to adjust its weights in its synapses to attempt to produce the output pattern. Afterwards, validation and testing is performed to select the best set of weights, and to measure the accuracy of the model. The ability of the neural network lies in the number of neurons in the network, the interconnected patterns or topology between the neurons, and the weights of the synapses between the neurons of each layer, which is a function of the learning algorithm used during the training, [4, 10].

Two different types of neural networks have been used in this study: a feed forward neural network (FFNN), and a linear vector quantization (LVQ) network. A FFNN is a neural network with a simple design, typically having 3 layers—input layer, hidden layer, and output layer. The number of nodes in the input layer is equal to the number of variables within the data, while the output is the dependent variable to which the inputs are being mapped to. The number of nodes in the hidden layer, and the number of hidden layers are experimentally determined by the amount needed to solve the problem. There are however two rules that are often used as a starting point to determine these values: (i) most problems can be solved using one hidden layer (ii) the number of hidden neurons is the mean of the neurons in the input and output layers. The basic schematic of a FFNN with n inputs and m outputs is found below in Fig. 1 [4, 10].

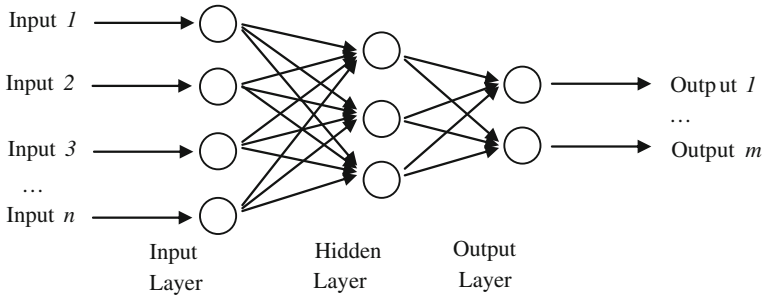


Fig. 1 Schematic of a feed forward neural network

A LVQ network utilises an unsupervised layer as its first hidden layer, followed by a linear transformation function, similar to a FFNN. The unsupervised layer, often called a self-organising map (SOM), allows data to be presented with no accompanying target information. Instead of classifying data based on predetermined classes, the SOM uses a similarity metric to cluster data with similar vectors in such a way that neurons which are physically near each other in the neuron layer respond to similar input vectors. Hence when the weights of a neuron are updated, the weights of its neighbouring neurons are updated as well. The topology of the SOM is usually in the shape of connected squares or hexagons. Figure 2 gives a schematic of a typical SOM with its topology in the shape of squares, 2 input units for each input vector, and 15 output units, [4, 20].

The second layer of the LVQ is a linear layer, which takes the cluster centres obtained in the SOM layer, and labels the clusters according to predetermined classes based on a majority of the points around the cluster centres that belong to the cluster. This class information is used to fine tune the location of the cluster centre in order to minimise the number of misclassified points. After this learning stage is done, the LVQ is able to classify an input vector by assigning it to the class, defined by the closest cluster centre, which its output unit in the SOM is closest to. Figure 3 shows the layout of a LVQ network, [4, 20].

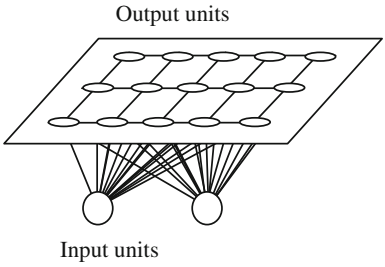


Fig. 2 SOM Network with *2* input units and *15* output units

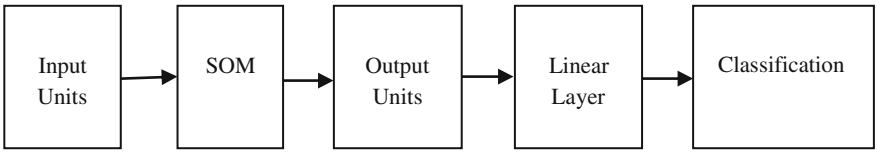


Fig. 3 Layout of a learning vector quantization (LVQ) network

4 Experiment

The overall goal of the experiment was to be able to classify a centrifugal pump as being either: (1) no cavitation (2) incipient cavitation or (3) full on cavitation. In order to do so, a three step procedure was proposed to pre-process the data obtained from an accelerometer signal found on the non-drive end bearing of a centrifugal pump, and then to determine which of the three categories the data falls into.

4.1 Methodology

The first step utilised key indicators of cavitation developed by McKee et al. McKee et al. utilised an altered form of octave band analysis and had determined that there were 3 octave bands that best show the existence of cavitation in an acceleration signal: octave band 2, 8 and 9, [8]. Since RMS velocity is used in ISO 10816 as a standard for determining the health of a centrifugal pump, the RMS velocity values from these three octave bands were obtained to characterise the data, [14, 15]. Using these as inputs, the second step performed a PCA on the RMS velocity values in these octave bands to determine three principal components that can best show the spread of the data. The last step takes the principal components as inputs into a neural network, either a FFNN or a LVQ, to characterise the data into one of three categories describing the state of the pump. For both the FFNN and the LVQ, the 170 data points from the pumps were randomized and then divided into 3 sets: 70 % of the data was placed in the training set, 15 % of the data was placed in the validation set, and 15 % of the data was placed in the testing set.

The size of the FFNN varied from having 1 hidden layer with 2 neurons to 2 hidden layers with 20 neurons each. The inputs to the FFNN were the RMS velocity values from the 3 octave bands, and the output was a number ranging from 0 to 1. The activation function for both the hidden layers and the output layers was a tangent-sigmoidal function, also known as ‘atan’, to limit the range of the output of the neurons. The training function was chosen to be the Levenberg-Marquardt method. To train, validate, and test the data, inputs were given with target outputs to the FFNN, and the mean squared error was used to evaluate the outputs of the FFNN. To determine the effectiveness of the FFNN, the amount of deviation of the results from the target data using the testing data was analysed. Three ranges of

accuracy were analysed: within 1 % deviation from the target value, within 10 % deviation of the target value, and within 50 % deviation from the target value.

The size of the SOM layer in the LVQ varied from 2 to 30 neurons. The mean squared error was used to evaluate the outputs of the LVQ. Unlike the FFNN where the output of a neuron of the FFNN may range from −1 to 1, the output of the neuron in the output layer is only a 1 or 0 in a category. As a result, there are no levels of correct classification, but only if the point was classified correctly or not.

4.2 Data

Three sets of data obtained from multiple sources were utilised for this experiment. In order to train the FFNN and the LVQ, data was required from a pump in each of the 3 categories. In addition, in order to account for a wide variety of pump characteristics, data from different sized pumps were obtained. Table 1 lists the pumps used in this experiment, pump condition, and the ISO 10816 values associated with the points. It has been found that the ISO 10816 levels are not always able to reflect the existence and severity of cavitation in a centrifugal pump. This is due to the fact that cavitation excites high frequency components in the vibration, which may have low amplitudes at incipient cavitation. Since the ISO 10816 severity charts evaluate the RMS velocity of the centrifugal pump, these low amplitudes may not alter the RMS velocity values enough to produce noticeable changes. As a result, centrifugal pumps with some amount of cavitation may be wrongly classified as non-faulty and containing a no cavitation.

From each pump, a series of 1 min vibration time data sets were analysed to determine the characteristics of the condition of the pump. Out of the 170 points used as data from the pumps, 81 points (47.65 %) were considered points from a non-cavitating pump, 30 points (17.65 %) were considered points representing incipient cavitation, and 59 points (34.7 %) were considered points representing cavitation, [8].

Table 1 Pumps used in experiment

Condition of pump	Type of pump
Non-cavitating	1. Gould 3,700, 90 kW, 4 vane impeller, 2,990 RPM
ISO 10816—Level A	2. Thompsons, Kelly and Lewis 400 × 450 ECSD 3 stage pump, LSE 2,600 kW, 1,485 RPM
Incipient cavitation	1. Gould 3,700, 90 kW, 4 vane impeller, 2,990 RPM
ISO 10816—Level A, B	2. Thompsons, Kelly and Lewis 400 × 450 ECSD 3 stage pump, LSE 2,600 kW, 1,485 RPM
Full-onset of cavitation	1. Thompsons, Kelly and Lewis 33″/36″ SDS-DV, 840 kW, 741 RPM
ISO 10816—Level B, C and D	2. Thompsons, Kelly and Lewis 24″/27″ SDS-DV, 446 kW, 986 RPM
	3. ITT Flygt CT-3231, 2 Blade, 85 kW, 1,475 RPM

5 Discussion

The sizes for both types of neural networks were changed in order to determine the ideal network size which would produce the highest percentage of classification. Sizes for a FFNN ranged from a single hidden layer with 2 neurons in the layer to a double hidden layer with 20 neurons in each layer. The number of neurons in the input and output layer were kept constant at 3 each, since this was determined by the number of octave bands used and the categories for the vibration. Likewise, the number of neurons for the competitive layer of the LVQ ranged from 2 to 30 neurons. The highlights of the results are in Table 2, which shows the percentage of points correctly classified from the test set of data. Each of these values is an average of the results of the ANN over a series of 10 trials.

Table 2 shows a number of results from the FFNN, which differ according to levels of accuracy and size of the neural network. A 1 % deviation of the target value means that the value in the neuron for the target category (no cavitation, incipient cavitation, full-onset of cavitation) was found to be 0.99 and higher. Likewise, a 10 % or a 50 % deviation of the target value means that the value in the neuron for the target category was found to be 0.9 and higher or 0.5 and higher. These different levels are stated since the output neurons in the FFNN produce levels of certainty of the classification of a point in the target category. Since LVQ output neurons produce only 0 or 1, the percentage classification is the same for all three evaluations of the neural network. Percentage values for these categories were obtained by dividing the number of points out of the test set that met the criteria by the total number of points in the test set.

The best FFNN with a single hidden layer had 7 neurons in its hidden layer and produced a 88.46 % success rate at the 10 % level, which is the least accurate of the four FFNN highlighted in Table 2. Using the rules of thumb in Sect. 3.3 to determine the number of neurons in the hidden layers, it was found that the optimal neural network fitting this description had 3 neurons in its first hidden layer and 1 in its second hidden layer, producing a 93.54 % success rate at the 10 % level. This neural network was found to the best at classification at the 1 % level. The neural network with the largest success rate at the 10 % level, which was 98.08 %, had the

Table 2 Highlights of percent classified correctly

Type of neural network	Number of neurons in hidden layers	Percentage classified correctly within 1 % of target value (%)	Percentage classified correctly within 10 % of target value (%)	Percentage classified correctly within 50 % of target value (%)
FFNN	7,0	83.46	88.46	89.23
FFNN	3,1	92.31	93.54	94.23
FFNN	19,16	90.38	98.08	98.85
FFNN	20,10	70.00	95.00	99.62
LVQ	10	56.54	56.54	56.54

size of 19 neurons in its first hidden layer, and 16 in its second. However, this was not reflective as being the best at the 1 % level, nor at the 50 % level. Finally, the FFNN that had the best classification at the 50 % level was found to have 20 neurons in the first layer and 10 neurons in the second layer.

Comparing the results of the neural networks, if this were to be deployed into industrial applications to detect cavitation, it would be recommended to use the neural network with 3–1 configuration in the hidden layers since it utilises the least amount of resources in terms of computation time and power. If resources are not a problem, then a FFNN of 19–16 would be ideal to use due to its high success rate at the 10 and 50 % level.

The LVQ, on the other hand, has shown to be very ineffective compared to the FFNN. Its peak performance was at 56.54 %, using 10 neurons in its competitive layer. This large difference in ability is due to the role of the unsupervised learning in the classification of points to certain clusters. It seems from the results, that the LVQ cannot completely separate the clusters of data, especially between the non-cavitating data and the incipient cavitation. As a result, the LVQ misclassifies almost 43 % of the points due to their proximity to another class' centre.

6 Conclusion

The results of this experiment show the plausibility of using a ANN to classify the state of cavitation in a centrifugal pump. Based on the percentage of correctly classified data in the test data set, the FFNN has outperformed the LVQ in its ability to correctly classify the condition of a centrifugal pump as having either a non-cavitating state, incipient cavitation state, or full-onset of cavitation state. This is due to the lack of distinct clusters to separate the three conditions. In the case of the FFNN, at a level of correct classification within 10 % of its target value, it was found that a FFNN with 19-16 hidden layers was optimal to use. However, in cases where resources are limited, a FFNN with 3-1 hidden layers was optimal to use due to its 93.54 % success rate at the 10 % level, and due to it having the highest success rate, 92.31 %, at the 1 % level.

The application of this method in industry would provide users with a quick and simple method of evaluating the amount of cavitation in their pumps. If automated, this method would be able to take a one minute vibration reading from an accelerometer on a centrifugal pump, and return to the user an integer from 1 to 3 which gives the state of cavitation in the pump. For a technician that is obtaining vibration readings periodically on a pump, this would allow the technician to determine the state of cavitation without having to record the vibration history of the pump. Instead, within minutes, the state of cavitation would be easily known and the technician would be able to take action accordingly.

References

1. Abdi H, Williams LJ (2010) Principal component analysis. Wiley Interdiscip Rev Comput Stat 2(4):433–459. doi:10.1002/wics.101
2. Amirat Y, Benbouzid MEH, Al-Ahmar E, Bensaker B, Turri S (2009) A brief status on condition monitoring and fault diagnosis in wind energy conversion systems. Renew Sustain Energy Rev 13(9):2629–2636. doi:http://dx.doi.org/10.1016/j.rser.2009.06.031
3. Bin L, Yaoyu L, Xin W, Yang Z (2009) A review of recent advances in wind turbine condition monitoring and fault diagnosis. In: Power Electronics and Machines in Wind Applications, PEMWA 24–26 June, 2009. IEEE, pp 1–7. doi:10.1109/pemwa.2009.5208325
4. Hameed Z, Hong YS, Cho YM, Ahn SH, Song CK (2009) Condition monitoring and fault detection of wind turbines and related algorithms: A review. Renew Sustain Energy Rev 13 (1):1–39. doi:http://dx.doi.org/10.1016/j.rser.2007.05.008
5. Klema J, Flek O, Kout J, Novakova L (2005) Intelligent diagnosis and learning in centrifugal pumps. In: Emerging solutions for future manufacturing systems. Springer, New York, pp 513–522
6. McKee K, Forbes G, Mazhar I, Entwistle R, Howard I (2013) A review of machinery diagnostics and prognostics implemented on a centrifugal pump. In: Jay Lee JN, Jag Sarangapani, Joseph Mathew (ed) Proceedings of the 6th World Congress on Engineering Asset Management, Springer, Cincinnati, OH, USA, Oct 2 2011
7. McKee KK, Forbes G, Mazhar I, Entwistle R, Howard I (2012a) Modification of the ISO-10816 centrifugal pump vibration severity charts for use with Octave band spectral measurements. In: 7th Australasian Congress on Applied Mechanics, Adelaide, SA, Dec 9–12 2012. Engineers Australia, pp 276–283
8. McKee KK, Forbes G, Mazhar I, Entwistle R, Hodkiewicz M, Howard I (2012b) A single cavitation indicator based on statistical parameters for a centrifugal pump. In: World Congress on Engineering Asset Management, Daejon, South Korea
9. Palgrave R (2005) Centrifugal pump basics. World Pumps 2005(460):37–39
10. Peck JP (1994) On-line condition monitoring of rotating equipment using neural networks. ISA Trans 33(2):159–164
11. Rayner R (1995) Pump users handbook. Elsevier Advanced Technology Oxford
12. Rencher AC, Christensen WF (2012) Principal component analysis. In: Methods of multivariate analysis. Wiley, West Sussex, pp 405–433. doi:10.1002/9781118391686.ch12
13. Standardization IOf (1975) ISO 532: Method for calculating loudness level. International Organization for Standardization, Switzerland
14. Standardization IOf (1998) ISO 10816-3: Mechanical vibration—Evaluation of machine vibration by measurements on non-rotating parts—Part 3: Industrial machines with nominal power above 15 kW and nominal speeds between 120 r/min and 15,000 r/min when measured in situ ISO, Switzerland
15. Standardization IOf (2009) ISO 10816-7: Mechanical vibration-Evaluation of machine vibration by measurements on non-rotating parts. Part 7: Rotodynamic pumps for industrial applications, including measurements on rotating shafts. ISO, Switzerland
16. Wang H (2010) Intelligent diagnosis methods for plant machinery. Front Mech Eng China 5 (1):118–124
17. Wang H, Chen P (2007) Sequential condition diagnosis for centrifugal pump system using fuzzy neural network. Neural Inf Process Lett Rev 11(3):41–50
18. Wang H, Chen P (2009) Intelligent diagnosis method for a centrifugal pump using features of vibration signals. Neural Comput Appl 18(4):397–405. doi:10.1007/s00521-008-0192-4
19. Wang Y, Liu Hou L, Yuan Shou Q, Tan Ming G, Wang K (2009) Prediction research on cavitation performance for centrifugal pumps. In: IEEE International Conference on Intelligent Computing and Intelligent Systems, ICIS 20–22 Nov 2009, pp 137–140. doi:10.1109/icicisys.2009.5357921

20. Wenxian Y, Tavner PJ, Crabtree CJ, Wilkinson M (2010) Cost-effective condition monitoring for wind turbines. IEEE Trans Indust Electr 57(1):263–271. doi:10.1109/tie.2009.2032202
21. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. Chemom Intell Lab Syst 2(1–3):37–52. doi:http://dx.doi.org/10.1016/0169-7439(87)80084-9
22. Zouari R, Sieg-Zieba S, Sidahmed M (2004) Fault detection system for centrifugal pumps using neural networks and neuro-fuzzy techniques. In: Paper presented at the Surveillance 5 CETIM Senlis 11–13 Oct 2004

Remaining Useful Life Estimation of Slurry Pumps Using the Health Status Probability Estimation Provided by Support Vector Machine

Peter W. Tse and Changqing Shen

Abstract The slurry pump is one of the most important machines and widely applied in oil sand industries, mining, waste treatment, etc. The mixtures transported by the pumps include the solids as well as the liquids with different volume and hardness that make the pumps work under abrasive and erosive environment. This would cause the continuous wear of the components, especially the impeller, in the pumps. As a result, the efficiency and useful life will be greatly reduced. Every unexpected failure of slurry pumps could cost companies high up to millions of dollars. To avoid this problem, traditional scheduled maintenance strategies are usually adopted but it can not warn the impending failure and sometimes the components are replaced when they are still in healthy status. Consequently, effective condition monitoring and online health status assessment methods are of great significance and should be developed to conduct timely and effective slurry pump fault diagnosis and prognosis. In this paper, an effective data driven technique for estimating the remaining useful life of slurry pumps are developed based on the health status probability estimation obtained by Support Vector Machines (SVMs). The signals collected by the sensors installed on an industrial slurry pump are used for analysis. The results show that frequency band selection and the position of sensors have some effect on the useful information acquisition and that SVM has superior performance in industrial data processing.

1 Introduction

As a key equipment to transport the mixture of solids and liquids, the slurry pump is widely used in the oil companies. The abrasive and erosive solid particles go through the components like impellers in the pump and this will accelerate the wear

P.W. Tse (✉) · C. Shen
Department of Systems Engineering and Engineering Management, City University of Hong Kong, Kowloon Tong, Hong Kong
e-mail: meptse@cityu.edu.hk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_9

of the pump [1]. Even worse, the impending failure and the degradation couldn't be effectively warned by traditional scheduled period maintenance strategies. As a result, much loss is caused by sudden break down. For this reason, both the pump manufactures and companies are pursuing potential methods to conduct effective fault detection. From the above, aiming for a reliable, safe, economical and efficient running of the slurry pump, the online health status identification is a very important and significant task.

Recently, some research work on revealing the health status of the slurry pump has been conducted. Walker et al. [2] discussed the relationship between the wear and the solid particle sizes, slurry concentration and pump speeds. Bross et al. [3] proposed a new model which could be used to predict the influence of the different impeller design parameters on the wear behaviour. In order to clarify the failure mechanisms, a systematic study on the failing impeller of a slurry pump used in wet process phosphoric acid manufacture has been carried out by means of scanning electron microscopy [4]. The results demonstrated that the weight loss of corrosive wear is affected by the impact velocity and its tangential and normal components in different areas of the impeller. The above achievement offer beneficial guidelines for the pump design and manufacturing process. However, in practical operation, the above methods may not be able to offer sufficient warning of the impending failure.

In order to simulate the practical working environment, some test rigs are established to carry out the case studies. Hancock et al. [5] decomposed vertical pump vibration signals from their hydraulic pump test stand using wavelet packet analysis, and then the packets containing signal features distinguishing normal and failed pump operation were entered into an adaptive neuro-fuzzy inference system (ANFIS) for pump health classification. Wang et al. [6] designed an experimental system for the slurry pump and some working conditions were adjustable in order to study their effect on the wear process of the impeller. Meanwhile, some data analysis based on this test rig has been conducted. Qu et al. [7] developed a data cleaning algorithm based on support vector machine and random sub-sampling validation followed by feature selection. The tests showed good capability of the data cleaning algorithm in identifying outliers for all datasets. Qu et al. [8] proposed a pump prognostics algorithm jointly uses least square support vector regression (LSSVR), genetic-algorithm-based optimization, and cumulative sum (CUSUM) technique. Although the above research made satisfied achievement, their performance under practical industrial environment is unknown.

The artificial-intelligence-based fault diagnosis methods have the potential to carry out online pump health condition monitoring [9]. These methods aim to recognize different pump health status via the features extracted from the vibration signals. The accuracy of the identification of these conditions can be further enhanced through classifiers that exhibit good performance [10]. Vanpnik [11] recently developed supervised learning approaches known as SVM based on statistical learning theory. SVM has well-defined formulations and keep consistent with mathematical theory. Moreover, owing to their superior generalization capability, SVMs do not require a large amount of samples for training [12]. Hu et al.

[13] used SVM ensembles for bearing fault diagnosis and achieved better results compared with other approaches.

In this paper, an effective data driven technique for estimating the practical remaining useful life of slurry pumps are developed based on the health status probability estimation obtained by SVMs. The signals collected by the sensors installed on an industrial slurry pump are used for analyses. Some spectral energy features are extracted from the raw signals under different wear conditions which have known remaining useful life. An intelligent pump health status probability estimation model is then constructed via the one-against-one SVMs. The result shows that the positions of sensors have some effect on the useful information acquisition and that support vector machine has satisfactory performance in industrial data processing. The rest of this paper is outlined as follow. Section 2 briefly describes the fundamental theory of SVMs. The proposed pump prognostics method is presented in Sect. 3, followed by industrial case verification as stated in Sect. 4. Some factors are discussed in this section. Conclusions are drawn in Sect. 5.

2 Theoretical Background of SVMs

The SVM classifier is a supervised learning algorithm based on statistical learning theory developed by Vapnik [11]. This technique maps the low dimensional data sets to the high dimensional feature space, and aims to solve a binary problem by searching an optimal hyper plane which can separate two data sets which have the largest margin in the high dimensional space. The optimal hyper plane is then established through a set of support vectors from the original data sets and these subsets form the boundary between the two classes. Given $\{x_i, y_i\}_{i=1}^{N}$ be a training data set where x_i is the input feature vector for each sample; N is the sample number; and $y \in \{-1, +1\}$ represents its label. As shown in Fig. 1, the optimal hyper plane is defined by $w \cdot x + b = 0$, where x is the point lying in the hyper plane, w is the parameter for the orientation of hyper plane, and b is a scalar

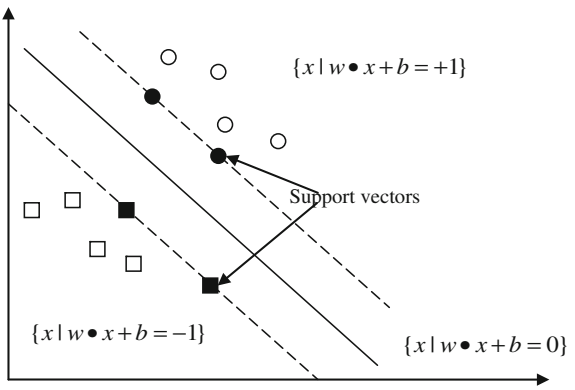


Fig. 1 Data automatically classified by SVM

threshold which represents the bias from the margins. For both classes, the input feature vectors satisfy the following inequality:

$$y_i(w \cdot x_i + b) \geq 1, i = 1, 2, \ldots, N \tag{1}$$

Hence, a decision function that can be used to decide whether a given data belongs to either the positive class or the negative class is given as follows:

$$f(x) = sign(w \cdot x + b) \tag{2}$$

The positive slack variable ξ_i is introduced to the above optimization problem and then a new optimization problem is formed as follow:

$$\begin{aligned} &\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \\ &s.t. \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i, i = 1, 2, \ldots, N \end{aligned} \tag{3}$$

where C is a penalty parameter which controls the tradeoff between the margin maximization and error minimization. After solving the Lagrange equation of Eq. (3), a classification function can be defined as:

$$f(x) = \text{sgn}\left\{\sum_{i=1}^{n}\alpha_i y_i K(x_i, x) + b\right\} \tag{4}$$

where α_i is the Lagrange multiplier; $K(x_i, x) = \varphi(x_i) \cdot \varphi(x)$ is a symmetric positive defined kernel function given by the Mercer's theorem, and the kernel function can map a low dimensional vector to a high feature space through some nonlinear functions. In this paper, the popular radial basis function (RBF) is adopted and its mathematical formula is given as:

$$K(x_i, x) = \exp\left(\frac{-\|x_i - x\|^2}{2\sigma^2}\right) \tag{5}$$

where σ is a positive real number.

The SVM is actually a binary classifier. However, the slurry pump usually experience different more than two wear stages. To solve this problem, a number of multiclass classification strategies, such as the one-against-one (OAO), the one-against-all (OAA) and the direct acyclic graph (DAG) could be employed [14]. The OAO based strategy is a voting approach which is regarded as a more suitable strategy to the actual application due to its comparatively fast training speed and good classification accuracy [15]. As shown in Fig. 2 for a k class problem, the OAO method constructs $k(k-1)/2$ binary classifiers. If x is classified to class I, the vote for class I is increased by one. Hence, the class with most votes is determined as the class for x and the OAO based strategy is also called Max Win strategy.

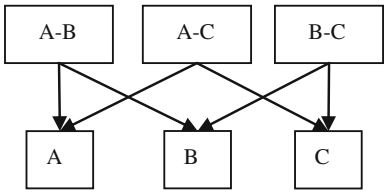


Fig. 2 The illustration of OAO approach

3 The Proposed Slurry Pump Remaining Useful Life Estimation Method

As a key component in the slurry pump, the continuous operation of the impeller is of great importance to ensure the production efficiency of the slurry pump. Hence, effective slurry pump impeller remaining useful life estimation could enhance the reliability of the slurry pump and prevent the unexpected breakdown. The slurry pump is a very complicated machine under the industrial environment. The fault related impacts could be easily whelmed by the noises from different sources. Hence, it is very hard to detect the impeller wear condition through signal processing based methods although they have been proved to be effective in identifying the rotating machinery health condition. Considering that the vibration energy of the impeller varies with the wear degree. The spectrum energies in sensitive frequency band are extracted as indicators. The OAO SVMs are used for pump remaining useful life estimation. Figure 3 depicts the procedure of the proposed method.

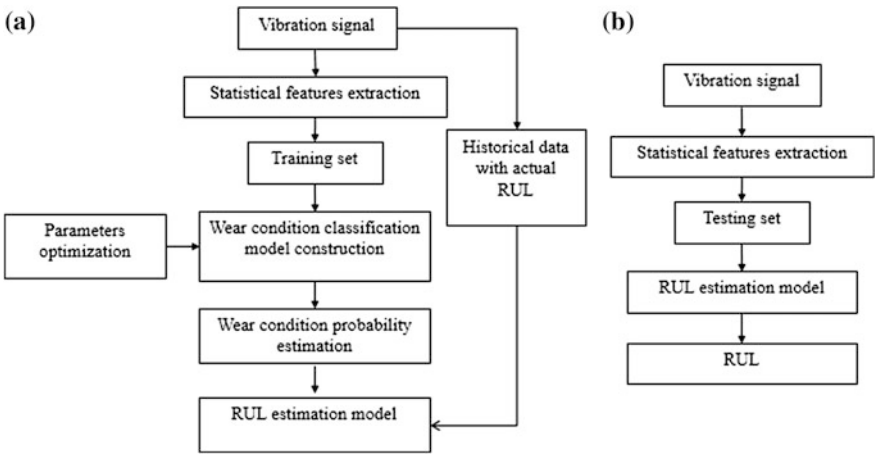


Fig. 3 The framework of the proposed pump RUL estimation method

The spectrum is divided into a number of sub-bands and then the spectrum energies are calculated from each sub-band. The vibration signal collected under different pump wear stage exhibit different spectrum energies in sensitive frequency bands. As a result, these statistical features are extracted for the further analysis. During the running of the slurry pump, the impeller will go through different wear stages and its remaining useful life decreases with the time. Hence, recognition of the pump health status is of great significance due to its strong relationship with the remaining useful life. Here, the OAO SVMs are employed to train and test the pump health status recognition model based on the historical data whose wear condition and remaining useful life are known.

Assume that there are K data sets which indicate K wear conditions with increasing degrees and decreasing RUL, then a K classes SVM multi-classifier is established. For a unknown dataset including J samples collected within a workday, each sample is fed to the classifier to judge its wear stage y_j ($y_j = 1, 2, 3, \ldots K, j = 1, 2, 3, \ldots J$) until the results for all the samples within a workday are obtained. From the above SVM multi-classification results, the probabilities of each health status are presented as following:

$$\begin{aligned} \mathrm{P}(s = k|x_1, x_2, \ldots, x_J) &= \sum_{j=1}^{J} \frac{I_k(y_j)}{J} \\ I_k(y) &= \begin{cases} 0 & y \neq k \\ 1 & y = k \end{cases}, k = 1, 2, \ldots, K \end{aligned} \tag{6}$$

The sum of the probabilities for each health status follows the following equation:

$$\sum_{k=1}^{K} P(s = k|x_1, x_2, \ldots, x_J) = 1 \tag{7}$$

After the estimation of current pump health status in term of probability distribution of each health status, the RUL estimation could be performed. Combining each health status probabilities for a dataset and the historical remaining useful life, the RUL estimation of the slurry pump can be expressed as Eq. (8) accordingly:

$$\mathrm{RUL} = \sum_{k=1}^{K} P(s = k|x_1, x_2, \ldots, x_J) \cdot \tau_k \tag{8}$$

where τ_k is the historical remaining useful life for the kth wear condition.

4 Industrial Case Studies

To validate the applicability of the proposed method, the data collected from the industrial slurry pump are analysed here. Four accelerometers were mounted at different positions, where the PCB 352A60 accelerometers (S1 and S2) were mounted on 'Casing Lower' and 'Casing Discharge' and the PCB 352C18 accelerometers (S3 and S4) were mounted on the 'suction and discharge pipes'. These four accelerometers collected the vibration signals from four different positions at a similar sampling frequency rate of 50 kHz. The pump vibration measurements were collected via the smart asset management system (SAMS). The data acquisition equipment consisting of a National Instrument (NI) cDAQ 9172 and a DAQ module NI 9234 was used. The data were recorded from March to June and the impeller was broken at the end of May according the historical data. The data sets listed in Table 1 are analysed here.

Different frequency bands are analysed here to investigate their sensitivities, the frequency band is divided into 8 sub-bands and thus 8 energy indicators are calculated for each sample. 10 samples of each data set are used for wear condition recognition model training and the rest 10 samples are used for testing. Once the pump health status probabilities for each sample are obtained. The estimated RUL could be calculated via Eq. (8). Some frequency bands in low frequency region are investigated. Figures 4 and 5 show the training and testing results for the slurry pump RUL estimation based on the data collected from S1. In this analysis, the frequency band is selected as 0–1,000 Hz and the width of each sub-band is 125 Hz. From the results, it could be seen that the training results matches well with the pump actual RUL. Although the trend of the results for the testing samples agree with the actual RUL, the deviations between the predicted RUL and the actual RUL for each data set are not satisfactory.

Table 1 Data used for testing purpose

Month	Date	RUL (Days)	Number of samples
March	13th	76	20
	20th	69	20
April	2nd	56	20
	9th	49	20
	16th	42	20
	21st	37	20
	28th	30	20
May	2nd	26	20
	22nd	6	20
	28th	0	20

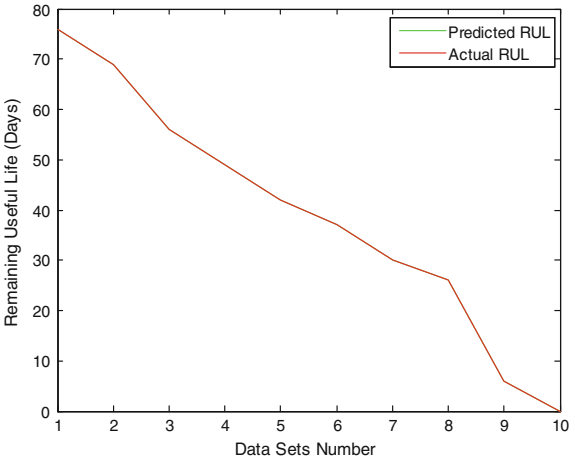


Fig. 4 The predicted RUL for the training samples whose features are extracted from 0–1,000 Hz

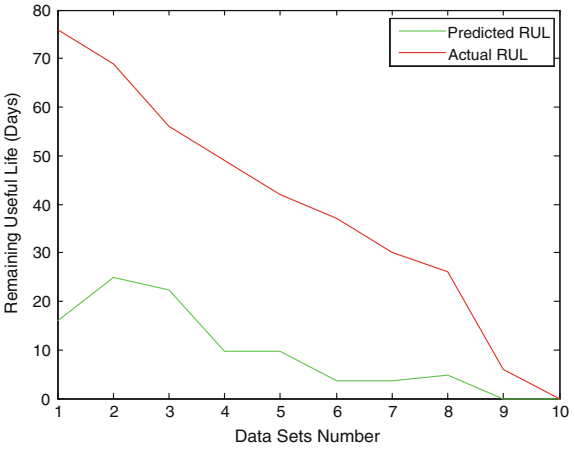


Fig. 5 The predicted RUL for the testing samples whose features are extracted from 0–1,000 Hz

The frequency band is then selected as 0–400 Hz and the width of each sub-band is 50 Hz. Similarly, Figs. 6 and 7 present the RUL prediction results for training and testing samples. It could be seen that the new frequency band is more sensitive for better results are obtained for the testing samples as shown in Fig. 7. The predicted RUL for the training samples keep in good accordance with the actual RUL. While some deviations between the predicted RUL and the actual RUL exists in the results for the testing samples. However, the deviations are smaller than those in Fig. 5. From the above analysis, the selection of sensitive frequency band for features extraction is an essential step to ensure satisfactory performance of the RUL estimation method. Moreover, as mentioned before, there are 4 sensors installed at

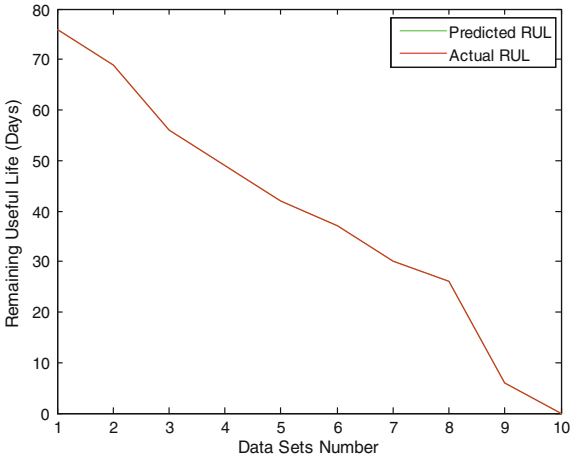


Fig. 6 The predicted RUL for the training samples whose features are extracted from 0–400 Hz

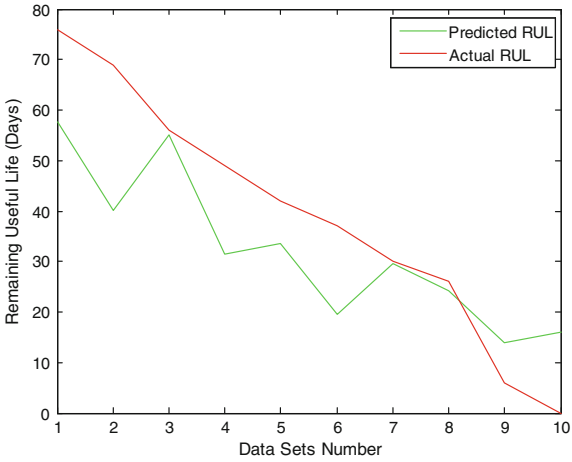


Fig. 7 The predicted RUL for the testing samples whose features are extracted from 0–400 Hz

different positions and thus the sensitivities of the sensor locations should also been discussed. Accordingly, the data from the other 3 sensors are processed as the above procedures. Figures 8, 9 and 10 present the predicted RUL for testing samples collected from the other 3 locations.

The following indicator is used for evaluating the performance of the data from different sensors:

$$EE = \sum_{k=1}^{K} |RUL_k - \tau_k| \tag{9}$$

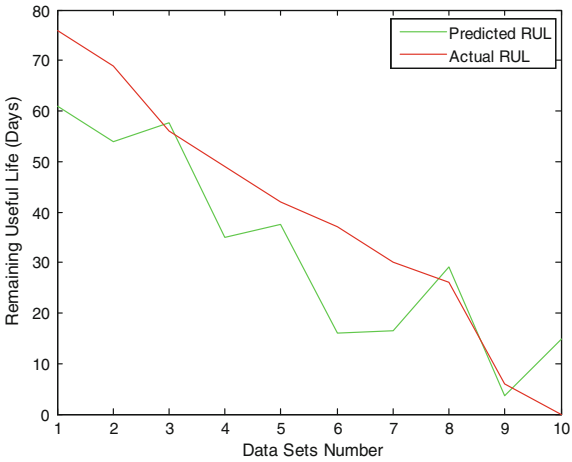


Fig. 8 The predicted RUL for the testing samples collected via S2

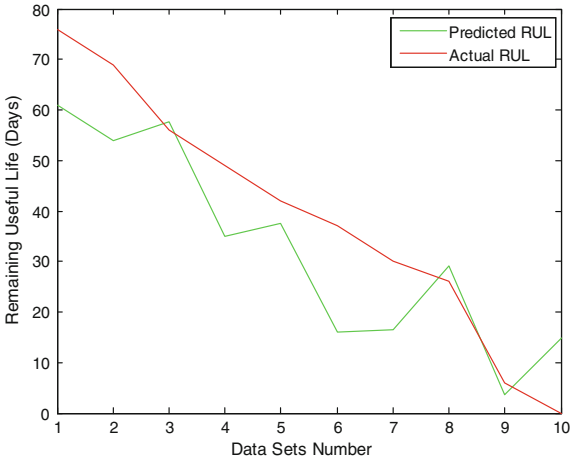


Fig. 9 The predicted RUL for the testing samples collected via S3

The indicator values for these 4 sensors are shown in Table 2:

According to the results in Table 2, the positions of the sensors also affect the performance of the analysis. The data from the sensors mounted on 'Casing Lower' and 'Casing Discharge' performs better than those from the other two places.

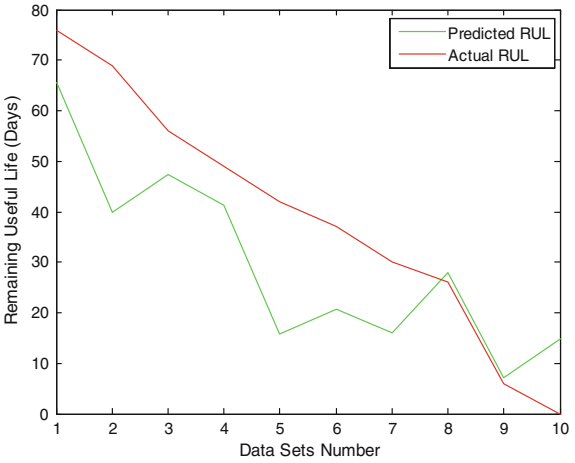


Fig. 10 The predicted RUL for the testing samples collected via S4

Table 2 The fault indicators calculated by each sensor

Sensor No.	1	2	3	4
Indicator value	117.6	105.1	130.3	130.3

5 Conclusion

In this paper, an new data driven technique for estimating the practical remaining useful life of slurry pumps are developed based on the health state probability estimation obtained by OAO SVMs. The spectral energy features are extracted from the raw vibration signals under different wear conditions which have known remaining useful life. An intelligent pump health status probability estimation model is then constructed via the one-against-one SVMs. Different frequency bands are investigated and the performances of the signals from different sensors are compared. The results prove that the range of 0–400 Hz is sensitive to the pump health status and the sensors installed at the 'Casing Lower' and 'Casing Discharge' could capture the pump status information more accurately.

Acknowledgment This article was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 122011) and a grant from City University of Hong Kong (Project No. 7008187).

References

1. Li P, Cai Q, Wei B (2006) Failure analysis of the impeller of slurry pump used in zinc hydrometallurgy process. Eng Fail Anal 13:876–885
2. Walker CI, Bodkin GC (2000) Empirical wear relationships for centrifugal slurry pumps: Part 1: side-liners. Wear 242:140–146
3. Bross S, Addie G (2002) Prediction of impeller nose wear behaviour in centrifugal slurry pumps. Exp Therm Fluid Sci 26:841–849
4. Fan A, Long J, Tao Z (1995) Failure analysis of the impeller of a slurry pump subjected to corrosive wear. Wear 181–183:876–882
5. Hancock KM, Zhang Q (2006) Hybrid approach to hydraulic vane pump condition monitoring and fault detection. Trans Am Soc Agric Biol Eng 9:1203–1211
6. Wang Y, Zuo MJ, Fan X (2005) Design of an experimental system for wear assessment of slurry pumps. In: Proceedings of the Canadian Engineering Education Association, Canada, pp 1–7
7. Qu J, Zuo MJ (2010) Support vector machine based data processing algorithm for wear degree classification of slurry pump systems. Measurement 43:781–791
8. Qu J, Zuo MJ (2012) An LSSVR-based algorithm for online system condition prognostics. Expert Syst Appl 39:6089–6102
9. Caesarendra W, Widodo A, Yang BS (2011) Combination of probability approach and support vector machine towards machine health prognostics. Probabilist Eng Mech 26:165–173
10. Di Maio F, Hu J, Tse P et al (2011) Ensemble-approaches for clustering health status of oil sand pumps. Expert Syst Appl 39:4847–4859
11. Vapnik VN (1995) The nature of statistical learning theory. Springer, Berlin
12. Shen CQ, Wang D, Kong FR et al (2013) Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier. Measurement 46:1551–1564
13. Hu Q, He ZJ, Zhang ZS et al (2007) Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMS ensemble. Mech Syst Signal Pr 21:688–705
14. Yang BS, Han T, Hwang WW (2005) Fault diagnosis of rotating machinery based on multi-class support vector machines. J Mech Sci Technol 19:846–859
15. Zhao SL, Zhang YC (2008) SVM classifier based fault diagnosis of the satellite attitude control system. In: International Conference on Intelligent Computation Technology and Automation, 2008, pp 907–911

Bearing Defect Diagnosis by Stochastic Resonance Based on Woods-Saxon Potential

Siliang Lu, Qingbo He and Fanrang Kong

Abstract The interference from background noise makes it difficult to identify the incipient bearing defect via vibration analysis. Stochastic resonance (SR) is a nonlinear phenomenon which is characterized by that the output signal can be enhanced with the assistance of the proper noise. The Large Parameter Bistable SR (LPBSR) method is commonly used in the bearing fault diagnosis. However, the LPBSR method requires signal tuning to satisfy the small parameter requirement (both amplitude and frequency are far less than 1), which implies the inherent structure of the input signal is modified and the effect of SR may be further affected. Additionally, a barrier exists in the bistable model, which indicates that the particle motion in the bistable potential is unstable and then the external noises induced by the unstable motion are introduced in the output signal. This chapter proposes a new strategy to realize bearing defect diagnosis, that is, a Woods-Saxon potential instead of a conventional bistable potential is utilized to achieve SR and to enhance the output signal-to-noise ratio (SNR). In the proposed method, the output SNR can be optimized just by tuning the parameters of the Woods-Saxon potential. This method overcomes the limitation of the small parameter requirement of the classical bistable SR, and can thus detect a high driving frequency. Furthermore, the smooth Woods-Saxon potential leads to a more stable particle motion compared to the bistable potential, which provides a more regular output waveform and reduces the unexpected noises at the same time. The proposed method has yielded more effective results than the traditional methods, which was verified by means of a practical bearing vibration signal carrying defect information.

S. Lu · Q. He (✉) · F. Kong
Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei 230026, Anhui, People's Republic of China
e-mail: qbhe@ustc.edu.cn

S. Lu
e-mail: lusliang@mail.ustc.edu.cn

F. Kong
e-mail: kongfr@ustc.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_10

1 Introduction

Rolling bearing is widely used in the industrial applications. The inner-race, outer-race along with the rolling elements in the bearing endure the radial alternating stress when a bearing rotates. Under this circumstance, the bearing may be defective over a period of time, a slight defect may cause the machine stop but a serious defect can cause a catastrophe. Hence, the condition maintenance and the fault diagnosis of the bearing are of significant. Traditional bearing fault defect diagnosis methods include vibrational and acoustical signal analyses. For the vibrational signal analysis method, the vibration sensor attached to the machine is utilized to measure the vibration signals of the bearing. And then the signal processing methods (e.g. time-domain signal analysis, frequency-domain signal analysis) are employed to analyse the measured signals. Based on the principle that the features of the normal bearing signal are different from that of the defect bearing signals, the health condition of the bearing can be determined after the signal analysis.

However, the bearing is just one of the components of a rotating machine, the vibrational signal from the bearing is not isolate, and it is always submerged in the background noises induced by other coupled components' vibration. To get a clean bearing signal, signal filtering is always in demand. Traditional signal filtering methods generally focus on suppressing the noise to improve the signal-to-noise ratio (SNR) of the output. For example, a weak signal feature extraction method based on wavelet transform (WT) is suggested in Ref. [1]. A method of using higher order cumulants (HOCs) and WT to improve the fault classification accuracy under poor SNR is proposed in Ref. [2]. Moreover, in Ref. [3], a wavelet-based approach for denoising coarsely quantized signals is proposed. In addition, the removal of discrete frequency noise using self-adaptive noise cancellation is also demonstrated for analysing the vibration signals [4]. These noise removal techniques are manipulated in frequency-domain, but it is noteworthy that the noise frequencies may be close to the bearing defective frequency. Consequently, the noise suppression may weaken the amplitude of the defective frequency at the same time, which may thus cause possible mistake in bearing defect diagnosis.

Distinct from the traditional signal denoising methods, stochastic resonance (SR) can utilize the proper noises to enhance the output signal, and has been manifested to be an effective method for weak signal detection [5]. In Ref. [6], a method for detecting signals buried in noise via nanowire transistors using SR is proposed. In Ref. [7], weak signal detection in a circuitry system is conducted with multiscale noise tuning SR. And in Ref. [8], an application of parameter-induced SR to target detection in shallow-water reverberation is introduced. The classical SR theory is based on the small parameter assumption (both amplitude and frequency of the signal are far less than 1); however, the practical signals from engineer applications are always with amplitude and frequency being higher than 1. To address this issue, large parameter bistable SR (LPBSR) is suggested [9]. By the LPBSR method, the signal is tuned to satisfy the small parameter requirement of SR. However, the signal tuning may change the inherent structure of the signal, and then the effect of

SR may be further affected. Besides, a barrier exists in the bistable potential, which indicates that the particle motion in the bistable potential is unstable and then the external noises induced by the unstable motion are introduced in the output signal.

Since the SR system can be regarded as a signal-input-output-system, the output signal can be affected by both the input signal and system structure. In this chapter, we investigate a new method based on Woods-Saxon potential SR to realize bearing defect diagnosis. The proposed method requires tuning SR system parameters rather than tuning signal parameters, and it is suitable for large parameter signal. Different from traditional SR system parameter tuning methods based on the bistable potential model, a new potential called the Woods-Saxon potential is proposed for SR system tuning in weak signal detection. The Woods-Saxon potential provides a smooth bottom, which indicates that the particle motion in the potential is stable. As a result, the output waveform from Woods-Saxon model is more regular than that from bistable model. In this study, experimental investigation by a practical defective bearing signal is addressed to verify the effectiveness of the proposed method for detecting the bearing's defective frequency in comparison with the envelope spectrum analysis and traditional LPBSR methods.

2 Traditional Large Parameter Bistable Stochastic Resonance

The traditional SR has been developed in bistable system. The SR phenomenon can be briefly described as: a particle is driven by the periodic force and the random noise in a bistable potential, and the periodic motion can be enhanced by the proper noise. Considering an overdamped case, the governing equation of the particle motion in the bistable potential can be written as:

$$\frac{dx}{dt} = -U'(x) + A_0 \sin(2\pi f_0 t + \varphi) + n(t) \tag{1}$$

where A_0 is the periodic signal amplitude, f_0 is the driving frequency, φ is the phase. $U(x)$ represents a reflection-symmetric quartic potential:

$$U(x) = -\frac{1}{2}ax^2 + \frac{1}{4}bx^4 \tag{2}$$

in which a and b are barrier parameters with positive real value. And $n(t) = \sqrt{2D}\xi(t)$, with $<n(t)n(t+\tau)> = 2D\delta(t)$ is the noise item, wherein D is the noise intensity and $\xi(t)$ represents a Gaussian white noise with zero mean and unit variance. Then Eq. (1) can be written as:

$$\frac{dx}{dt} = ax - bx^3 + A_0 \sin(2\pi f_0 t + \varphi) + \sqrt{2D}\xi(t) \tag{3}$$

The most important feature of the bistable SR is that the output amplitude depends on the noise intensity D [5]. Specifically, the output amplitude firstly increases with increasing noise level, reaches a maximum, and then decreases again, which indicates the celebrated SR effect. Via the SR, the output signal can be optimized by adjusting the noise level.

The above theoretical analysis based on bistable model is under the assumption that the amplitudes and frequencies of the input signals (periodic signal and noises) are all smaller than 1. This can be interpreted by the adiabatic approximation or linear response theory. To make the SR suitable for processing the large parameter signal, the signal should be tuned based on normalized scale transformation [10]. Mathematically, let $y = x\sqrt{b/a}$, $\tau = ax$ and $K = \sqrt{a^3/b}$, and substitute them to Eq. (3), we can obtain:

$$\frac{dy}{d\tau} = y - y^3 + \frac{1}{K}\left[A_0 \sin\left(\frac{2\pi f_0 \tau}{a} + \varphi\right) + \sqrt{2D}\xi\left(\frac{\tau}{a}\right)\right] \tag{4}$$

Equation (4) is the normalized form of Eq. (3). From Eq. (4), it can be seen that the driving frequency has been normalized to be $1/a$ times of the original frequency. Consequently, the large parameter SR can be achieved by applying Eq. (4). With the normalized scale transformation, choosing a large parameter K can normalize a high frequency (larger than 1 Hz) to be much smaller than one, which hence satisfies the requirement of the classical SR.

3 Stochastic Resonance Based on Woods-Saxon Potential

3.1 Basic Principle

From the aforementioned analyses, it can be found that the output signal is relative to the interaction between the driving signal and the random noises. In fact, the system output can be also affected by the SR potential $U(x)$ [11]. In this study, the capacity of weak signal detection based on Woods-Saxon potential SR is investigated. The Woods-Saxon potential is a mean-filed potential for single nucleon states within the shell model of nuclear structure that firstly proposed by Woods and Saxon [12]. And Ignacio Deza et al. used the Woods-Saxon potential for a study of wide spectrum energy harvesting based on the SR principle [13]. In this study, we intend to extend the application of Woods-Saxon SR in signal processing region. The equation of Woods-Saxon potential is shown below:

$$U(x) = -\frac{V_0}{1+\exp\left(\frac{|x|-r}{c}\right)} \tag{5}$$

Equation (5) demonstrates that the shape of Woods-Saxon potential $U(x)$ is determined by three parameters, where parameter V_0 affects the depth of the potential, parameter r affects the width of the potential while parameter c determines the smoother of the potential. An intuitional illustration of $U(x)$ at fixed $V_0 = 2$, $r = 1$ and varying c from 0.01 to 0.28 is shown in Fig. 1.

It can be found in Fig. 1, $U(x)$ becomes a square well of depth V_0 and width $2r$ for $c = 0$. As c grows the potential walls become smoother, maintaining the value $-V_0/2$ at $|x| = r$. When c is large exceedingly, $U(x)$ resembles a parabolic potential. Hence the potential shape can be adjusted by parameters V_0, r and c respectively. Substitute Eq. (5) to Eq. (1), the equation of SR based on Woods-Saxon potential can be obtained as:

$$\frac{dx}{dt} = \frac{V_0}{c}\mathrm{sgn}(x)\exp\left(\frac{|x|-r}{c}\right)\left(1+\exp\left(\frac{|x|-r}{c}\right)\right)^{-2} + A_0\sin(2\pi f_0 t+\varphi)+n(t) \tag{6}$$

where $\mathrm{sgn}(x)$ denotes the sign function. It can be seen from Eq. (6), the right hand side of Eq. (6) is consisted of three items: the first item related to the potential structure, the second item related to the periodic signal and the third item related to the noise. And the system output which shown in the left hand side of Eq. (6) can be adjusted by tuning the SR system parameters. This is the basis of Woods-Saxon SR.

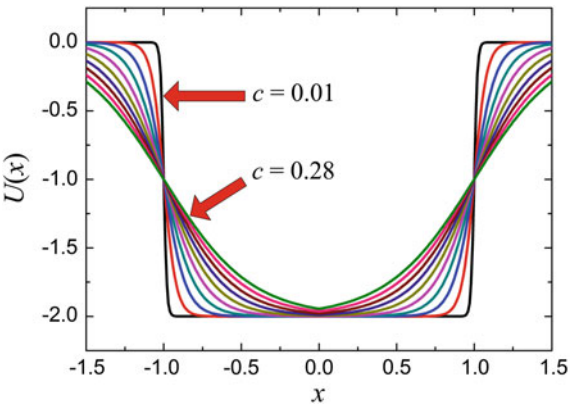


Fig. 1 Woods-Saxon potential $U(x)$ at fixed $V_0 = 2$, $r = 1$ and varying c from 0.01 to 0.28

3.2 Proposed Bearing Defect Diagnosis Scheme

When a fault is occurring in a bearing, periodic impulses can be revealed in the corresponding spectrum of the generated vibrational signals. Different types of bearing faults will lead to impulses appearing at different periodic intervals, thus the fault characteristic frequency is commonly analyzed to classify the fault type of the bearing [14]. For a quantitative analysis of the defective frequency detection effect, the SNR of the output signal is introduced as a criterion. The SNR is defined as follow:

$$SNR = 10\log_{10}\left(\frac{A_d}{A_n}\right) \tag{7}$$

wherein A_d and A_n represent the amplitudes corresponding to the driving frequency f_d and the strongest interference frequency f_n (which means the frequency with the largest amplitude except the driving frequency) in the power spectrum, respectively. A larger SNR indicates a better discrimination between the periodic signal and the noise. Based on such a criterion, the bearing defect diagnosis scheme based on Woods-Saxon SR can be illustrated in Fig. 2. The detailed steps of the algorithm are depicted as follows:

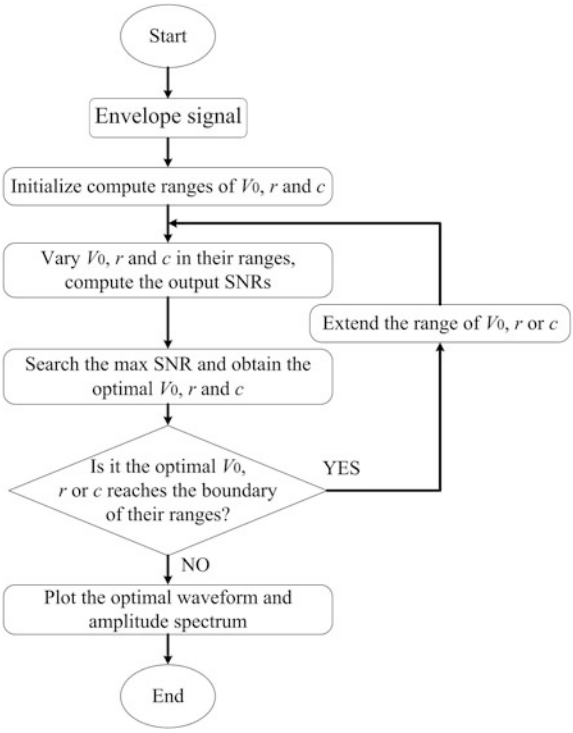


Fig. 2 The proposed bearing defect diagnosis scheme

With the above steps, the Woods-Saxon potential is adjusted to match the input signal and then the optimal output signal with maximal SNR can be obtained.

4 Experimental Verification

To verify the effectiveness of the proposed bearing defect diagnosis method, the defective bearing vibrational data from Case Western Reserve University (CWRU) Bearing Data Center Website [15] are utilized for analysis. The proposed algorithm is also compared to the traditional envelope and traditional LPBSR spectrum analysis methods.

As shown in Fig. 3, the test stand consists of a 2 hp motor (left), a torque transducer/encoder (center), a dynamometer (right), and control electronics (not shown). The motor shaft is supported by the test bearings. Vibrational signal was collected with the accelerometers attached to the housing with magnetic bases, and the sampling frequency is 12 kHz for drive end bearing experiments. The bearing used in this study is the deep groove ball bearing with the type of 6205-2RS JEM SKF. The detailed geometries of this bearing are shown in Table 1. In this study, the outer race defective bearing signal with rotation speed 1,723 rpm is analyzed. Single point fault was introduced to the test bearing's outer race using electro-discharge machining with fault diameters 14 mils (1 mil = 0.001 inches), and the defect depth is 11 mils. Based on the bearing's geometries and the rotation speed f_r, the ball passing frequency over the defect, f_{BPFO}, can be calculated as follow:

$$f_{BPFO} = \frac{1}{2}\left(1 - \frac{d}{D_m}\cos\alpha\right)\frac{f_n}{60}Z \tag{8}$$

where d, D_m, α and Z represent the ball diameter, the pitch diameter, the contact angle of the rolling element and the number of rolling elements, respectively. And then f_{BPFO} is calculated to be 102.5 Hz based on Eq. (8).

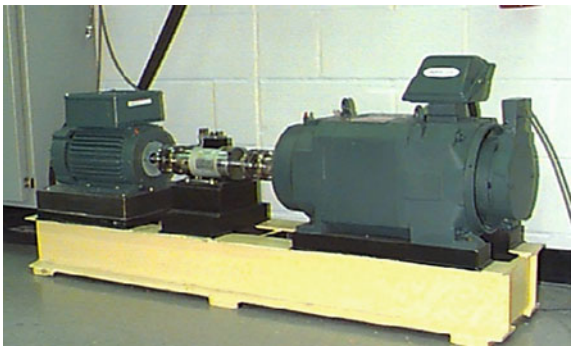


Fig. 3 The bearing test stand

Table 1 Geometries of the bearing (inches)

Outside diameter	Inside diameter	Thickness	Ball diameter	Pitch diameter
2.0472	0.9843	0.5906	0.3126	1.537

Firstly, the waveform and power spectrum of the original outer race defective signal are show in Fig. 4a. It can be found from the waveform that the impulses induced by the defect are submerged in the heavy background noise, and the f_{BPFO} component is also overwhelmed in the power spectrum. Considering that the machine resonance vibration are always demodulated by the defective impulses in the practical engineering applications, the Hilbert Transform (HT) method is introduced to demodulate the original signal and to reveal the substantial defective information [16]. The results that are processed by HT envelope method are shown in Fig. 4b, it can be seen from the power spectrum, the f_{BPFO} component can be pointed out by the arrow. However, the energies of the noise components are higher than that of the f_{BPFO}, which may cause possible error in bearing fault diagnosis. Besides, the SNR of the HT envelope signal is calculated to be −4.12 dB according to Eq. (7). In addition, the traditional LPBSR method that is introduced in Sect. 2 is applied to analyse the same signal, and the waveform and power spectrum are depicted in Fig. 4c with optimal parameter K = 360,000 according to Eq. (4). The power spectrum indicates that the f_{BPFO} is still hardly to be identified, and the corresponding SNR is just −3.50 dB. Such results are unsatisfied as the f_{BPFO} is still submerged in the surrounding noises.

Finally, the same outer race defective signal is processed by the proposed method following the steps shown in Fig. 2. It can be noticed from the power spectrum (Fig. 4d) with the optimal parameters V_0 = 8, r = 0.56 and c = 0.14, the f_{BPFO} is highlighted with the highest energy. And the other noise components are all with relative low energies as compared to f_{BPFO}, which makes the f_{BPFO} can be recognized at the first sight. The output signal's SNR based on the proposed Woods-Saxon SR method is 2.55 dB, which has a pronounced increase as compared to that from HT envelope and traditional LPBSR methods. From another aspect, it is noteworthy that, the peak amplitude of the time-domain waveform produced by the Woods-Saxon potential is regular as show in Fig. 4d. This can be interpreted by the shape of the Woods-Saxon potential shown in Fig. 1, the smooth potential bottom provides the particle's stable motion, and the potential wall steepness limits the particle motion in a specific range. As a result, the waveform produced by the Woods-Saxon potential is more regular than that from the bistable potential.

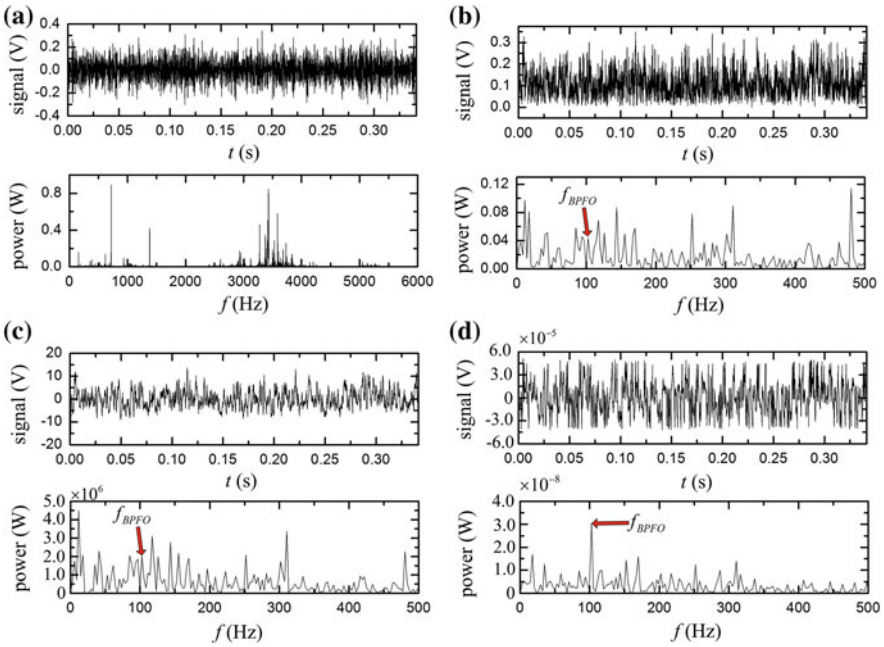


Fig. 4 Waveforms and power spectrums of: **a** the original signal; **b** the HT envelope signal; **c** the processed signal based on traditional LPBSR method (K = 360,000); **d** and the processed signal based on the proposed Woods-Saxon SR method (V_0 = 8, r = 0.56, c = 0.14)

5 Conclusions

A bearing fault diagnosis method based on Woods-Saxon potential SR has been investigated. In the proposed algorithm, the Woods-Saxon potential parameters are adjusted to match the input signal, which is different from the traditional LPBSR method (the signal is adjusted to satisfy the small parameter requirement). In this circumstance, the proposed method can detect bearing defective signal with large parameters while retaining the inherent information of the original signal. Besides, the smooth bottom together with the steep walls in Woods-Saxon potential provides a regular output waveform. The effectiveness of the proposed method has been manifested by the practical bearing signal carrying outer race defect in comparison with the HT envelope spectrum analysis and traditional LPBSR spectrum analysis methods.

Acknowledgment This work was supported by the National Natural Science Foundation of China (51075379, 51005221 and 11274300). The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

References

1. Wang CT, Gao RX (2003) Wavelet transform with spectral post-processing for enhanced feature extraction. IEEE Trans Instrum Meas 52(4):1296–1301
2. Yaqub MF, Gondal I, Kamruzzaman J (2012) Inchoate Fault detection framework: adaptive selection of wavelet nodes and cumulant orders. IEEE Trans Instrum Meas 61(3):685–695
3. Neville S, Dimopoulos N (2006) Wavelet denoising of coarsely quantized signals. IEEE Trans Instrum Meas 55(3):892–901
4. Ho D, Randall RB (2000) Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals. Mech Syst Signal Pr 14(5):763–788
5. Gammaitoni L, Hanggi P, Jung P, Marchesoni F (1998) Stochastic resonance. Rev Mod Phys 70(1):223–287
6. Nishiguchi K, Fujiwara A (2012) Detecting signals buried in noise via nanowire transistors using stochastic resonance. Appl Phys Lett 101(19):193108
7. Dai D, He Q (2012) Multiscale noise tuning stochastic resonance enhances weak signal detection in a circuitry system. Meas Sci Technol 23(11):115001
8. Xu B, Zhang H, Zeng L, Li J, Wu X, Jiang Z-P (2007) Application of parameter-induced stochastic resonance to target detection in shallow-water reverberation. Appl Phys Lett 91 (9):091908
9. Leng YG, Leng YS, Tai YW, Yan G (2006) Numerical analysis and engineering application of large parameter stochastic resonance. J Sound Vibrat 292(3–5):788–801
10. Yang DX, Hu NQ (2004) Detection of weak aperiodic shock, signal based on stochastic resonance. In: 3rd International Symposium on Instrumentation Science and Technology, Xi'an, China, 0210–0213
11. Lu S, He Q, Zhang H, Zhang S, Kong F (2013) Note: Signal amplification and filtering with a tristable stochastic resonance cantilever. Rev Sci Instrum 84(2):026110
12. Bohr A, Mottelson BR (1975) Nuclear structure, vol 1. Benjamin, New York
13. Deza JI, Deza RR, Wio HS (2012) Wide-spectrum energy harvesting out of colored Levy-like fluctuations, by monostable piezoelectric transducers. Europhys Lett 100(3):38001
14. Shen C, He Q, Kong F, Peter WT (2013) A fast and adaptive varying-scale morphological analysis method for rolling element bearing fault diagnosis. J Mech Eng Sci 227 (6):1362–1370
15. http://csegroups.case.edu/bearingdatacenter/pages/download-data-file
16. Peng ZK, Peter WT, Chu FL (2005) An improved Hilbert-Huang transform and its application in vibration signal analysis. J Sound Vibrat 286(1–2):187–205

Experimental Investigation on Suppressing Fluid-Induced Vibration in the Seal Clearance by Anti-swirl Flow

Chenglong Lv, Lidong He, Guo Chen, Peng Hu, Bingkang Zhang and Jinji Gao

Abstract The fluid-induced vibration caused by rotor eccentricity would increase the rotor vibration in the hydroelectric generating set, which would affect the security of rotating machines. The experiment rig is manufactured based on the simplified seal model. The influence of water pumping, water injection and gas injection to fluid-induced vibration in the seal clearance caused by rotor eccentricity is investigated by model experiment. The vibration amplitudes of the rotor with water pumping, water injection and gas injection at different positions and flows are contrasted. The result shows that the fluid-induced vibration in the seal clearance can be suppressed by the anti-swirl flow which has the right position and flow. And the best position and flow of the anti-swirl flow are found. It provides effective experimental basis for using anti-swirl flow technique to suppress the fluid-induced vibration in engineering applications.

Keywords Fluid-induced vibration · Water pumping · Water injection · Gas injection · Anti-swirl flow

1 Introduction

Vibration of hydroelectric generating set directly affects the stable operation of the unit and the economic benefits of hydropower station, and it also threatens the safety of hydropower station [1]. The fluid-induced vibration plays an important role in various factors which affect the stable operation of the hydro turbine. The cause of fluid-induced vibration is that the seal device of hydroelectric generating set is complex. The device has a small gap, and there is a big pressure difference

C. Lv · L. He (✉) · G. Chen · P. Hu · B. Zhang · J. Gao
Beijing University of Chemical Technology, Beijing, China
e-mail: he63@263.net

C. Lv
e-mail: lvchenglongsky@126.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_11

around the device. The pressure difference makes the rotor off-center and causes large pressure fluctuation that is the so-called fluid-induced vibration in the seal clearance [2]. Safety problems happen in operation because of the increase of the fluid-induced vibration.

The nature of the fluid-induced vibration is the oscillating force which is produced on the object. Vortex is a concrete manifestation of the oscillating force, and it is both a fine structure and a highly concentrated structure of the fluid kinetic energy [3]. The key of reducing the fluid-induced vibration is to destroy the generation of the vortex. Honeycomb seal is a passive control, which is an advanced technology in engineering by increasing the damping [4, 5]. Anti-swirl flow technology is an active control, which is a new technology [6–8]. The existing anti-swirl flow technology is that the fluid is injected directly into the sealed cavity. The vortex is destroyed by the fluid, so that the vibration can be reduced. The fluid-induced vibration and the unbalance response of the rotor seal system can be suppressed by the anti-swirl flow effectively, but the flow rate and flow velocity of the anti-swirl flow are not the bigger the better. The rotor vibration can be suppressed if the flow rate and flow velocity are appropriate, or the unstable vibration will be caused [9].

He et al. [10] and Yin [11] have researched that the rotor vibration could be suppressed by water pumping in the upstream position of the minimum clearance, but the best position of the water pumping is not found. The fact that the rotor vibration can be suppressed by the anti-swirl flow with the appropriate position and flow rate is proved by doing model experiment in this paper. And the best position and flow rate of the anti-swirl flow are found by measuring and contrasting the rotor vibration.

2 The Experiment Seal Model and Devices

2.1 The Experiment Seal Model

The fluid-induced vibration is caused by rotor eccentricity in the sealed cavity.

The purpose of this experiment is to prove that the fluid-induced vibration in the seal clearance can be suppressed by the anti-swirl flow. The anti-swirl flow can make circumferential pressure balance and reduce rotor vibration. The seal model is simplified as shown in Fig. 1.

The seal structure is composed of rotor and fixed ring, and there is a seal gap between them. When the rotor rotates, the fluid flow rotates with a high circumferential speed. The value of the rotor eccentricity is changed by adjusting the size of the gap between the rotor and the fixed ring. The anti-swirl flow experiment rig with single injector is manufactured based on the simplified seal model in this paper.

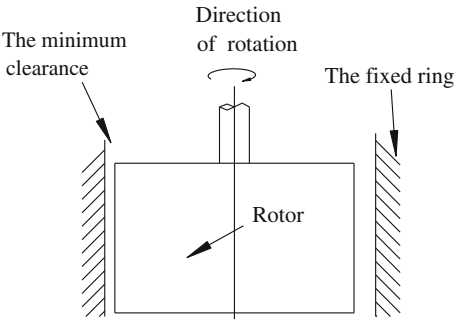


Fig. 1 The simplified seal model

2.2 Experiment Devices

Two stainless steel thin-walled cylinders are selected as the fixed ring and rotor which compose the sealing system. The larger diameter one is the fixed ring, the other one is the rotor which is driven by a motor. There is a radius gap of 6 mm between them when the fixed ring and rotor are axis concentric. A hole is punched on the fixed ring, and the nozzle is inserted along the inner wall of the fixed ring. The nozzle and the fixed ring are glued, and the fixed ring and the container are also glued. The water is added into the container, and make the water's height be same as the fixed ring's. The rotor rotates with the motor starting, and the rotational speed of motor is adjusted to 350 rpm. The distance between the rotor and the fixed ring is adjusted to 3 mm. The water is injected or pumped by starting water pump, and the air is injected by starting air pump. The rotor vibration is measured by eddy current sensor and OR38 data acquisition and analysis system. The specific experiment devices are shown as Fig. 2.

3 Water Pumping Experiment

The fluid-induced vibration occurred in the uneven seal clearance which is caused by the rotor eccentricity. The circumferential pressure distribute unevenly because of the fluid-induced vibration. The maximum dynamical pressure exists in the position of the minimum clearance. In order to ease pressure imbalance, water is pumped from the upstream position of the minimum clearance. The best position of water pumping is found by water pumping experiment.

Upstream position and downstream position of the minimum clearance are shown in Fig. 3. As shown, we set the position of the minimum clearance as 0°. 10° is the position that spinning 10 degrees towards upstream position of the minimum clearance. By that analogy, it has 20°, 30°, and so on. Similarly, −10° is the position that spinning 10 degrees towards downstream position of the minimum clearance, and it has −20°, −30°, and so on. Different degrees stand for different positions of upstream and downstream in this paper.

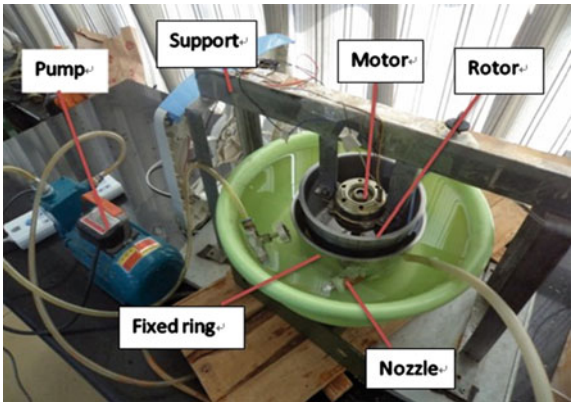


Fig. 2 Experiment devices

3.1 The Influence of the Position of Water Pumping to the Rotor Vibration

Experiment devices are connected according to the Fig. 2. The rotor vibration amplitude is 20 μm without water pumping. Then the water pump is started, and the valve is adjusted to make the flow rate value be 0.3 m^3/h. The positions of water pumping vary from −40° to 50°. The rotor vibration amplitudes of water pumping at different positions are recorded, as shown in Fig. 4.

As can be seen from the Fig. 4, the fluid-induced vibration is suppressed when the water is pumped at the upstream position of the minimum clearance. The rotor vibration amplitude decreases first then increases with the position of water pumping moving away from the position of the minimum clearance. The minimum amplitude of the rotor vibration is 16 μm at the position of 30°, and approximately 20 % vibration attenuation is obtained. The vibration amplitude increases with water pumping at the downstream position of the minimum clearance. The result shows that it has a negative effect when the water is pumped from the downstream.

4 Water Injection Experiment

4.1 The Influence of the Position of Water Injection to the Rotor Vibration

The import and export pipeline of the water pump are changed in order to inject the water into the seal clearance. The rotor vibration amplitude is 20 μm without water injection. Then the water pump is started, and the valve is adjusted to make the flow rate value be 0.3 m^3/h. The positions of water injection vary from −60° to 40°. The rotor vibration amplitudes of water injection at different positions are recorded, as shown in Fig. 5.

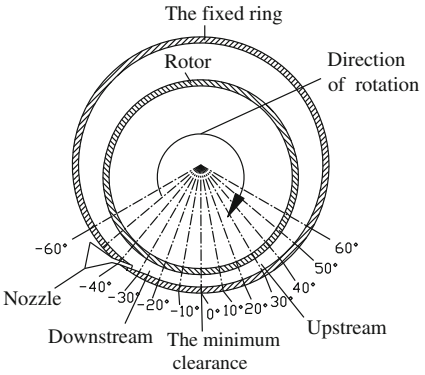


Fig. 3 Upstream position and downstream position of the minimum clearance

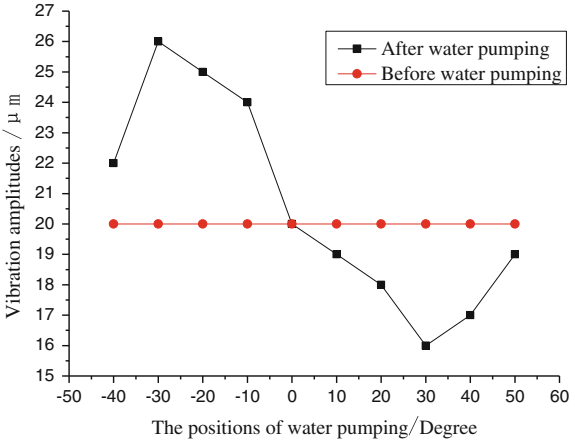


Fig. 4 The rotor vibration amplitudes of different locations before and after water pumping

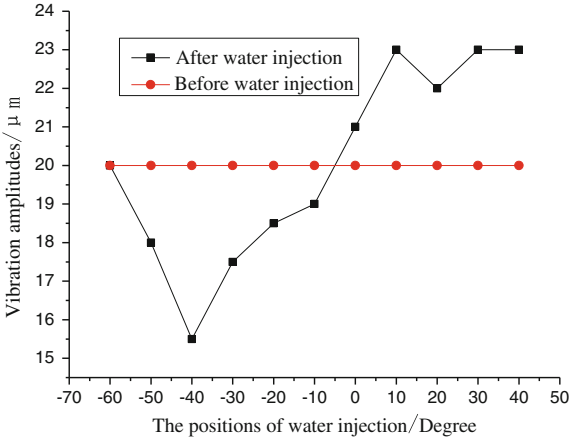


Fig. 5 The rotor vibration amplitudes of different locations before and after water injection

As can be seen from the Fig. 5, the fluid-induced vibration is suppressed when the water is injected at the downstream position of the minimum clearance. The rotor vibration amplitude decreases first then increases with the position of water injection varying from the position of −60° to the position of 40°. The minimum amplitude of the rotor vibration is 15.5 μm at the position of −40°, and approximately 22.5 % vibration attenuation is obtained. The Fig. 5 also shows that it has a negative effect when the water is injected at the upstream position.

4.2 *The Influence of the Flow Rate of Water Injection to the Rotor Vibration*

The rotor vibration amplitude is 20 μm without water injection. The injection position is set to the position of −40°. The valve is adjusted to make the flow rate value vary from 0.1 to 0.7 m^3/h. The rotor vibration amplitudes of different flow rate values are recorded, as shown in Fig. 6.

As can be seen from the Fig. 6, the rotor vibration amplitude decreases first then increases with the flow rate increasing. The best vibration attenuation effect is got when the flow rate value is 0.4 m^3/h. The minimum amplitude is 14 μm, and approximately 30 % vibration attenuation is obtained. The amplitude is greater than the initial vibration amplitude when the flow rate value exceeds 0.55 m^3/h. It shows that the flow rate of the anti-swirl flow is not the bigger the better for suppressing the fluid-induced vibration. It will have a negative effect if the flow rate is too big.

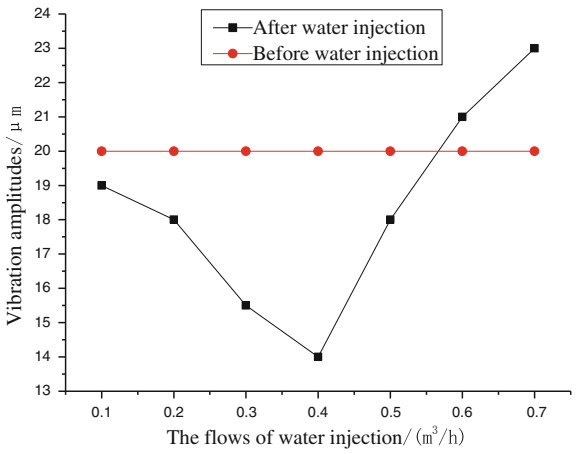


Fig. 6 The rotor vibration amplitudes of different flows before and after water injection

5 Air Injection Experiment

5.1 The Influence of the Position of Air Injection to the Rotor Vibration

The value of the rotor eccentricity and the rotational speed of motor are kept the same, and the water pump, liquid flowmeter are replaced by air pump, gas flowmeter. The rotor vibration amplitude is 25 μm without air injection which is slightly bigger than the former experiment because of the loss of water in the container. The valve is adjusted to make the flow rate value be 0.25 m^3/h, and the positions of air injection vary from −50° to 40°. The rotor vibration amplitudes of air injection at different positions are recorded, as shown in Fig. 7.

As can be seen from the Fig. 7, the rotor vibration amplitude decreases first then increases with the position of air injection varying from the position of −50° to the position of 40°. The minimum amplitude of the rotor vibration is 18 μm at the position of −30°, and approximately 28 % vibration attenuation is obtained. The Fig. 7 also shows that it has a negative effect when the air is injected at the upstream position.

5.2 The Influence of the Flow Rate of Air Injection to the Rotor Vibration

The measured rotor vibration amplitude is 27 μm without air injection. The injection position is set to the position of −30°. The valve is adjusted to make the flow rate value vary from 0.125 to 0.75 m^3/h, and the rotor vibration amplitudes of different flow rate values are recorded, as shown in Fig. 8.

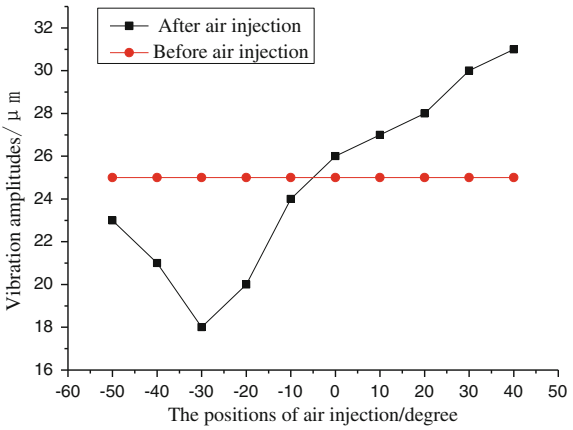


Fig. 7 The rotor vibration amplitudes of different locations before and after air injection

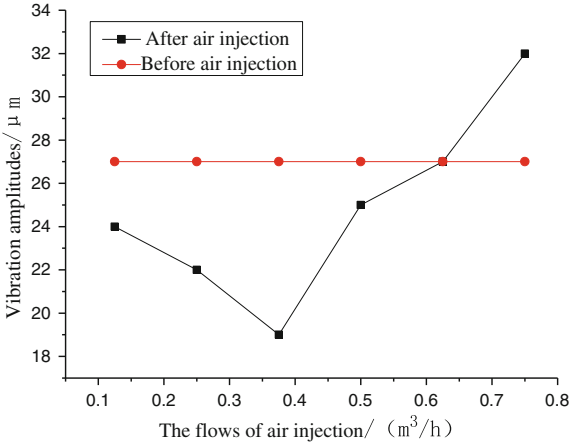


Fig. 8 The rotor vibration amplitude of different flows before and after air injection

As can be seen from the Fig. 8, the rotor vibration amplitude decreases first then increases with the flow rate increasing. The best vibration attenuation effect is got when the flow rate value is 0.375 m^3/h. The minimum amplitude is 19 μm, and approximately 29.6 % vibration attenuation is obtained. The amplitude is greater than the initial vibration amplitude when the flow rate value exceeds 0.625 m^3/h. The result shows that it will have a negative effect if the flow rate is too big.

6 Conclusions

The influence of water pumping, water injection and gas injection to the fluid-induced vibration in the seal clearance caused by rotor eccentricity is investigated by model experiment. We can get the following main conclusions:

(1) The fluid-induced vibration in the seal clearance can be suppressed when the water is pumped at the upstream position of the minimum clearance. The best vibration attenuation effect is got when the water is pumped at the position of 30°, and approximately 20 % vibration attenuation is obtained. It will have a negative effect if the water is pumped at any downstream position.

(2) The fluid-induced vibration in the seal clearance can be suppressed when the water is injected at the upstream position of the minimum clearance. The best vibration attenuation effect is got when the water is injected at the position of 40°, and approximately 22.5 % vibration attenuation is obtained. It will have a negative effect if the water is injected at any downstream position. The rotor vibration amplitude changes with the flow rate value of water injection. The rotor vibration amplitude is the smallest when the flow rate value is 0.4 m^3/h, and approximately 30 % vibration attenuation is obtained. It will have a negative effect if the flow rate value of water injection exceeds 0.55 m^3/h.

(3) The fluid-induced vibration in the seal clearance also can be suppressed when the air is injected at the upstream position of the minimum clearance. The best vibration attenuation effect is got when the air is injected at the position of 40°, and approximately 28 % vibration attenuation is obtained. It will have a negative effect if the air is injected at any downstream position. Like water injection, the rotor vibration amplitude changes with the flow rate value of air injection. The rotor vibration amplitude is the smallest when the flow rate value is 0.375 m^3/h, and approximately 20 % vibration attenuation is obtained. It will have a negative effect if the flow rate value of air injection exceeds 0.625 m^3/h.

Acknowledgements The authors gratefully acknowledge the financial support provided by Joint Project Special Fund of Education Committee of Beijing and the Ph. D. Programs Foundation of Ministry of Education of China (Item No.: 20110010110009) and National Basic Research Program of China (973 program) (Item No.: 2012CB026000).

References

1. Guo H (2005) The vibration research of mixed flow water turbine. Fujian Agric Mach 3:18–21
2. Yao D, Li Z, Qu D (1998) Large electric machine and hydraulic turbine. Anal Self-Excited Vib Francis-Turbine 5:43–47
3. Sun Y (2009) Design and experimental study on vibration reduction system using closed-loop control air-inhale technology. Beijing
4. Wang WZ, Meng G et al (2009) Nonlinear analysis of orbital motion of rotor subject to leakage air flow through an interlocking seal. J Fluids Struct 25(5):751–765
5. Dowson P, Walker MS, Watson AP (2004) Development of abradable and rub-tolerant seal materials for application in centrifugal compressors and steam turbines. Seal Technol 12:5–10
6. Shen Q, Li L, Pan Y (1994) Gas exciting force of labyrinth seal and its anti-swirl flow measure. Fluid Mach 22(7):7–12
7. Chen Y, Dong D (1994) Rotor vibration anti-swirl flow active control experimental study. J Aerosp Power 9(2):183–185
8. Merchant A, Kerrebrock JL, Epstein AH (2004) Compressors with aspirated flow control and counter-rotation. Am Inst Aeronaut Astronaut 2004–2514
9. He L (1999) The numerical simulation of suppressing vibration in rotor seal system. J Aerosp Power 14(3):1–7
10. He L, Yin D, Li C, Wu W (2009) Simulation experiment on water pumping and vibration suppression of Francis-turbine applying honeycomb seals. J Drain Irrig Mach Eng 28 (3):215–219
11. Yin D (2010) The vibration suppression study of water pumping and automatic controllable air pumping in sealing cavity. Beijing

A New Optimization Method for ECT Sensor Design

Nan Li and X.D. Yang

Abstract A new parameters optimization method of ECT sensor based on the orthogonal experimental design is presented. The sensor structure and sensor model are proposed. The evaluation criteria for sensor performance are introduced. The uniformity index of empty pipe and half pipe filled with engine oil are set as the optimization objectives. The experiments are set up based on multi-index orthogonal design. The optimizing parameters of the sensor structure include the number of electrodes, electrode width (corresponding to central angel) and radius of pipe wall. The typical sensitivity maps of the optimized sensor are presented at last. The experimental results indicate that the method can derive an evenly distributed sensitivity field. Compared with the full collocation method, the numbers of experiments is reduced by 66 %.

1 Introduction

ECT(Electrical Capacitance Tomography) is defined as a typical PT(Processed Tomography) technique for visualizing dielectric processes [1]. The advantages of ECT includes low cost, no radiation and non-invasive. The study for ECT has lasted for the past two decades, including small capacitance measuring [2, 3] and data acquisition circuit [4], image reconstruction algorithms [5], and industrial applications [6, 7] Moreover, many researches have been processed in ECT sensor design. Effect of Number of electrodes in ECT sensors on image quality is presented in Ref. [8], and influence of shielding arrangement on ECT sensors is

N. Li (✉)
IET Member, College of Mechanical Engineering and Applied Electronics Technology, Beijing University of Technology, Beijing, China
e-mail: nan.li@bjut.edu.cn

X.D. Yang
College of Mechanical Engineering and Applied Electronics Technology, Beijing University of Technology, Beijing, China

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_12

discussed in Ref. [9] Peng et al. [10]. Analyzed the impacts of the length of both the measurement electrodes and the guard electrodes on the sensitivity distribution, Alme et al. [11]. Illustrate the design criteria involved in radial and axial dimensions of ECT sensor electrodes, especially in sensor guards design. A square ECT sensor structure is presented in Reference [12]. Penetration depth, measurement sensitivity, dynamic range, signal strength, noise tolerance and imaging resolution are used to evaluate the sensor performance. The sensor parameters are kind of design trade-offs. Literature review shows that few papers discussed these trade-offs. Normally, the sensor optimization is processed under the consideration of single parameter changes.

In this paper, three main parameters including the number of electrodes, electrode width (corresponding to central angel) and radius of pipe wall are considered at the same time. The paper aims to figure out the priority of the sensor parameters and study the effect of these three parameters in an ECT sensor on the quality of reconstructed images.

2 ECT Sensor

An ECT sensor consists of several electrodes. The electrodes are evenly placed around the nonconductive vessel to ensure the electrical field lines penetrate through the whole detection area. Earthed outside screen and radial shielding are proposed to reduce the influence of the noise from circumstance.

In order to analysis the effects of numerous design parameters, the ECT sensor is needed to be modelled. Equation (1) is commonly used to determine the change in permittivity distribution $\Delta\varepsilon(x, y)$ from the change in measurement capacitance ΔC, and Eq. (2) is used to calculate a sensitivity map as part of a sensitivity matrix for image reconstruction.

$$\Delta C = \mathbf{J}\Delta\varepsilon \tag{1}$$

$$\mathbf{S}_{ij}(x, y) = -\frac{\int_{\Gamma} \nabla\varphi_i(x, y) \cdot \nabla\varphi_j(x, y) dxdy}{V_i \cdot V_j} \tag{2}$$

where $\mathbf{J}$ is sensitivity matrix, and $S_{ij}(x, y)$ is defined as the sensitivity between ith electrode and jth electrode at $\Gamma(x, y)$, $\varphi_i(x, y)$ is the potential distribution when ith electrode is applied V_i volts as an excitation and the other electrodes are set as detectors. Similarly, $\varphi_j(x, y)$ is the potential distribution when jth electrode is activated as excitation electrode with V_j volts on it and the other electrodes are set as detectors.

Usually, the sensitivity distribution in image area is heterogeneous. The sensitivity near the pipe wall and the excitation electrode is much higher than the sensitivity of the central area. For better imaging of the central area, the sensitivity distribution should be as uniform as possible. Therefore, Eq. (5) is applied to evaluate the optimization results.

$$\mathbf{S}_{i,j}^{\mu} = \frac{1}{m}\sum_{k=1}^{m}\mathbf{S}_{i,j}(e_k) \tag{3}$$

$$\mathbf{S}_{i,j}^{\sigma} = \sqrt{\frac{\sum_{k=1}^{m}\left(\mathbf{S}_{i,j}(e_k) - \mathbf{S}_{i,j}^{\mu}\right)^2}{m}} \tag{4}$$

$$P_{cv} = \frac{1}{N}\sum\left|\frac{\mathbf{S}_{i,j}^{\sigma}}{\mathbf{S}_{i,j}^{\mu}}\right| \tag{5}$$

where $S_{i,j}(e_k)$ is sensitivity in kth element (e_k), the image area is divided into m pixels, k = 1, 2, 3,..., m. $S_{i,j}^{\mu}$ is the mean value of sensitivity, and $S_{i,j}^{\sigma}$ represents the standard deviation of the sensitivity. P_{cv} named uniformity index is coefficient of variation of the sensitivity distribution. It can be used to evaluate homogenization of the sensitivity in the image area. P_{cv} is a partial small index, the P_{cv} smaller the sensitivity distribution of image field uniformity better.

3 Experimental Results and Analysis

Since there is no analytical equation can accurately describe the relationship between the optimization objective function and the ECT sensor structure, the practical experiment is the only way to determine the optimized parameters. Normally, the number of experiments required for optimization is large, but when using an orthogonal design method, it is significantly reduced. Compared with the factor alternate method [13], the orthogonal design method is simple and the experimental results are accurate and reliable. Compared with uniform design [14], the orthogonal design does not need to establish a complex regression forecast model to carry out the second experiment based on the experimental results. Therefore, the experiments were arranged using an orthogonal design in this paper.

The experimental factors and levels for the experiments are shown in Table 1. The potential number of experiments is determined according to $L_9(3^4)$ orthogonal design table. Compared with the full coordination method, the number of experiments required using orthogonal design is reduced by 66 %.

The orthogonal experimental program and results of $L_9(3^4)$ are shown in Table 2. The experimental results are indicated in Fig. 1.

Since P_{cv} is partial small index, according to Fig. 2, the optimal combination can be derived. The maximum range of P_e is 3.677, and it means the number of sensor electrode is the most important parameter, the effects of three sensor parameters on uniformity index P_e are N > R>W, and the same situation is appeared when sensing area is filled with Engine oil in half pipe. The maximum range of P_s is 3.460. The effects of three sensor parameters on uniformity index P_s are N > R>W, too.

Table 1 $L_9(3^4)$ orthogonal design table (Factors-levels)

Factors levels	The number of electrodes N	Electrode width (corresponding to central angel) W(°)	Radius of pipe wall R(mm)
Group1	8	8	40
Group2	12	12	60
Group3	16	16	80

Table 2 $L_9(3^4)$ orthogonal experiment program and results

No.	N	W	R	P_e^a	P_s^b
1	1	1	1	5.0235	6.6955
2	1	2	2	4.7894	7.5524
3	1	3	3	5.8165	9.0112
4	2	1	2	8.8020	10.8983
5	2	2	3	6.2371	10.0490
6	2	3	1	6.9103	8.1323
7	3	1	3	8.1693	12.2005
8	3	2	1	8.5898	9.6513
9	3	3	2	9.2426	12.4379
$\bar{K}_{pe1}$	7.753	9.931	8.160		
$\bar{K}_{pe2}$	9.693	9.084	10.296		
$\bar{K}_{pe3}$	11.430	9.860	10.420		
R_{pe}	3.677	0.847	2.260		
$\bar{K}_{ps1}$	5.210	7.332	6.844		
$\bar{K}_{ps2}$	7.316	6.542	7.611		
$\bar{K}_{ps3}$	8.670	7.323	7.739		
R_{ps}	3.460	0.790	0.870		

[a] P_e is the uniformity index when sensing field is empty

[b] P_s is the uniformity index when sensing field has *engine oil* distributed in half of pipe

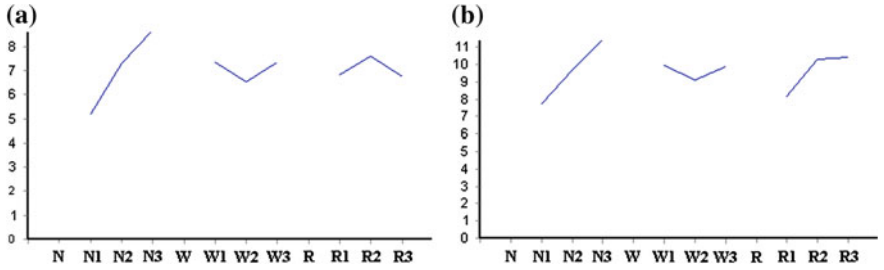


Fig. 1 Orthogonal experimental results **a** sensing area is empty; and **b** filled with Engine oil in half pipe

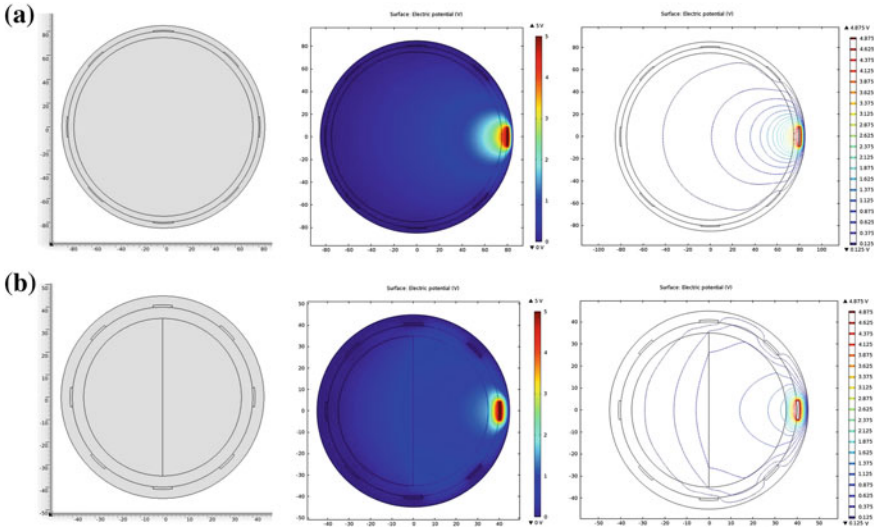


Fig. 2 The potential distribution in sensing field under different uniformity index P_e and P_s. **a** Simulation of the ECT sensor designed according to optimal combination N1W2R3; **b** Simulation of the ECT sensor designed according to optimal combination N1W2R1

Table 3 Optimal combination of sensor parameters

	P_e	P_s
Optimal combination	N1W2R3	N1W2R1
Uniformity index	4.3975	6.4855

The optimal combinations under P_e and P_s are shown in Table 3, respectively. According to the optimal combinations of the design parameters, the optimized sensor structure and the potential distribution in sensing field under different uniformity index P_e and P_s are presented in Fig. 2.

4 Conclusion

For typical non-linear systems, there are incompatibilities and conflicts amongst the parameters of the sensors. From the experiments in this paper, we can conclude that:

(1) It is inefficient to analysis the influence of the ECT sensor design parameters on sensor performance by using the full collocation method, since those many unknown factors which exist in practice at industry sites. The experimental design is important and can significantly reduce the number of trails, the

orthogonal design is applied in sensor parameter optimization experimental design.

(2) The optimal combinations of sensor design parameters are derived, which indicate the primary parameter is the number of sensor electrode, the secondary parameter is radius of the pipe all, and the last is electrode width. The sensitivity maps presented at the last prove the optimization of the sensor design parameters is effective and apparent.

Acknowledgments This work was partially supported by a project funded by National Natural Science Foundation of China (51105008), by Scientific Research Common Program of Beijing Municipal Commission of Education (KM201310005034), and supported by the Zhejiang Open Foundation of the Most Important Subjects.

References

1. Huang SM, Plaskowski AB, Xie CG, Beck MS (1989) Tomographic imaging of two-component flow using capacitance sensors. J Phys E: Sci Instrum 22:173–177
2. Yang WQ (1996) Hardware design of electrical capacitance tomography systems. Meas Sci Technol 7:225–232
3. Preethichandra DMG, Katsunori Shida (2001) A simple interface circuit to measure very small capacitance changes in capacitive sensors. IEEE Trans Instrum Meas 50:1583–1586
4. Liu S, Shen YJ, Zheng GB (2007) Design of data acquisition system for 12-electrode electrical capacitance tomography. In: Proceedings of the 2007 IEEE international conference on mechatronics and automation, Harbin
5. Yang WQ, Spink DM, York TA, Hccann H (1999) An image-reconstruction algorithm based on Landweber's iteration method for electrical-capacitance tomography. Meas Sci Technol 10:1065–1069
6. Warsito W, Fan LS (2001) Measurement of real-time flow structures in gas-liquid and gas-liquid-solid flow systems using electrical capacitance tomography (ECT). Chem Eng Sci 56:6455–6462
7. Wang HG, Wang XF, Lu QG, Sun YK, Yang WQ (2011) Imaging gas-solids distribution in cyclone inlet of circulating fluidised bed with rectangular ECT sensor, IEEE international conference on imaging systems and techniques (IST), Penang
8. Peng Lh, Ye Jm, Lu G, Yang WQ (2012) Evaluation of effect of number of electrodes in ECT sensors on image quality. IEEE Sens J 12:1554–1565
9. Olmos AM, Primicia JA, Marron JLF (2006) Influence of shielding arrangement on ECT sensors. Sensors 6:1118–1127
10. Peng LH, Mou CH, Yao DY, Zhang BF, Xiao DY (2005) Determination of the optimal axial length of the electrode in an electrical capacitance tomography sensor. Flow Meas Instrum 16:169–175
11. Alme KJ, Mylvaganam S (2006) Electrical capacitance tomography—sensor models. Des Simul Exp Verification, IEEE Sens J 6:1256–1266
12. Yang WQ, Liu S (1999) Electrical capacitance tomography with a square sensor. In: 1st world congress on industrial process tomography (WCIPT1), Buxton
13. Xie CG, Stott AL, Plaskowski A, Beck MS (1990) Design of capacitance electrodes for concentration measurement of two-phase flow. J Measur Sci Technol 1:65–28
14. Liang YZ, Fang KT, Xu QS (2001) Uniformdesign and its applications in chemistry and chemical engineering. J Chemom Intell Lab Syst 58:43–57

An Adaptive Doppler Effect Reduction Algorithm for Wayside Acoustic Defective Bearing Detector System

Fang Liu, Changqing Shen, Ao Zhang, Fanrang Kong and Yongbin Liu

Abstract In the wayside Acoustic Defective Bearing Detector (ADBD) system, because of the high moving speed of the railway vehicle, the recorded acoustic signal will be severely distorted by the Doppler effect, which is a barrier that would badly reduce the effectiveness of online defect detection. This paper proposes an adaptive Doppler effect reduction algorithm for the ADBD system. In this algorithm, firstly, the narrow-band signal is got by the band-pass filter after the sensitive frequency band selection; Secondly, the parameters of the Doppler kinematic model are estimated by maximizing the Pearson's correlation coefficient between the narrow-band signal and the Doppler atom; Finally, the Doppler-shifted signal is restored by the resampling method. The effectiveness of this method is verified by means of simulation studies and applications to diagnosis of train roller bearing defects.

1 Introduction

Roller bearing defect is the dominant type of fault for a train, which leads to serious accidents and significant costs for the rail transport industry [1]. Approximately 50 bearing related derailments occur in the United States each year [2]. So, it is important to develop techniques of condition monitoring and fault diagnosis system for the train bearings.

Wayside ADBD system [1] was developed in the 1980s to detect bearing defects of a moving train before overheated. The key technology of this system is based on the assumption that diagnostically relevant information is stored in the acoustic signal generated by the passing vehicle's bearings [2]. In comparison with the other systems, it costs lower and can detect bearing defects before overheated operation

F. Liu (✉) · C. Shen · A. Zhang · F. Kong · Y. Liu
Department of Precision Machinery and Precision Instrumentation,
University of Science and Technology of China, Hefei, China
e-mail: liufang1@mail.ustc.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_13

occurs or earlier in the failure process so that bearing maintenance can be performed on a scheduled basis [3].

In the ADBD system, the microphones are mounted by the wayside to collect acoustic signals emitted from the passing train bearings. As the high relative speed between the train and the microphone, the recorded signal is distorted by the Doppler effect which brings in the signal's frequency shift and frequency band expansion. So research on how to remove the signal's frequential structure disturbance is of great significance, especially for those methods based on frequency domain analysis.

In the 1990s, the Phase Locked Loop (PLL) method was proposed to correct the Doppler-shifted acoustic signal by Stojanovic et al. [4] Then the method which combined the PLL and Decision Feedback Equalize (DFE) algorithm was proposed by Johnson et al. [5] and it was applied to sonar communication between the autonomous underwater vehicle (AUV) and the surface ship. This technique was complicated as it was invented for the communication domain and many other techniques were involved. Recently, Dybała et al. [6] proposed a disturbance-oriented dynamic signal resampling method to correct the Doppler-shifted signal. In this method, the instantaneous frequency was acquired based on the Hilbert transformation, and then the Doppler-shifted signal is corrected via the resampling method. However, this method needs to know the characteristic frequency beforehand.

On the other hand, methods in the time domain have their special merits. Yang and Wang [7] established the time space relationship between the measurement field, the radiating field and the acoustic holography field, and then put forward a method based on nonlinear mapping function between the sound source and the measured signal, in which the Doppler effect is removed. But the geometric parameters must be known or be measured beforehand.

In this paper, an adaptive Doppler effect reduction algorithm for the ADBD system is introduced. In this algorithm, firstly, the narrow-band signal is obtained by the band-pass filter after the sensitive frequency band selection; Secondly, the parameters of the Doppler kinematic model are estimated by maximizing the Pearson's correlation coefficient between the narrow-band signal and the Doppler atom; Finally, the Doppler-shifted signal is restored by the resampling method. Compared with the methods based on instantaneous frequency estimation (IFE), this method does not need to know the characteristic frequency beforehand. Compared with the time domain methods, this method does not need to know the geometric parameters beforehand.

The rest of this paper is outlined as follows. Section 2 introduced the definition of Doppler atom. The correlation filtering analysis method is introduced in Sect. 3. The Doppler effect reduction method based on re-sampling method is briefly described in Sect. 4. The proposed adaptive Doppler effect reduction algorithm for the ADBD system is introduced in Sect. 5 in detail. An experimental verification test using defective train roller bearings with outer race defect is provided in Sect. 6. Finally, Sect. 7 draws concluding remarks.

2 Definition of Doppler Atom

In this paper, the basic model illustrated in Fig. 1 containing a single moving acoustic source and one signal receiver is considered. The source is moving along a straight line from point A to B with a constant speed.

Assume that the acoustic source of the train bearing with subsonic velocity is a monopole point source, and the medium has no viscosity and no energy loss, then Eq. (1) can be derived according to the wave equation and the moving relationship [8].

$$P = \frac{q'[t-(R/c)]}{4\pi R(1-M\cos\theta)^2} + \frac{q[t-(R/c)]\cdot(\cos\theta - M)V_0}{4\pi R^2(1-M\cos\theta)^3} \tag{1}$$

The P in the equation stands for the received sound pressure. q stands for the total quality flow rate of the source point, and $q' = \partial q/\partial t$. R denotes the distance between the source point and the microphone at the emitting time, c denotes the velocity of waves in the medium of air, θ stands for the angle between the vector of source point's velocity and the stretch between the source point and the microphone. V_O stands for the source point's velocity and $M = V_O/c$ stands for the Mach number of the source point's velocity.

The first part of Eq. (1) shows the inverse relation between the sound pressure and the distance between the source point and the microphone. The second part stands for the near-field effect. When the measurement is under the far-field condition or the Mach number is below 0.2, the near-field effect can be ignored [9], so the received sound pressure can be expressed as

$$P_{v_0,r,S,c,q} = \frac{q'[t-(R/c)]}{4\pi R(1-M\cos\theta)^2} \tag{2}$$

when the sound source is given as harmonic with the total quality flow rate of $q = q_O \sin(2\pi f_O t + \varphi)$, the received sound pressure will be

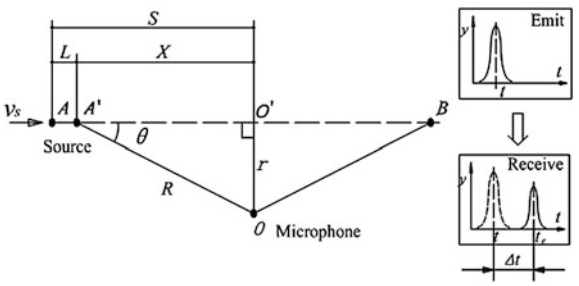


Fig. 1 Kinematics of radiation from a single acoustic source moving in a straight line

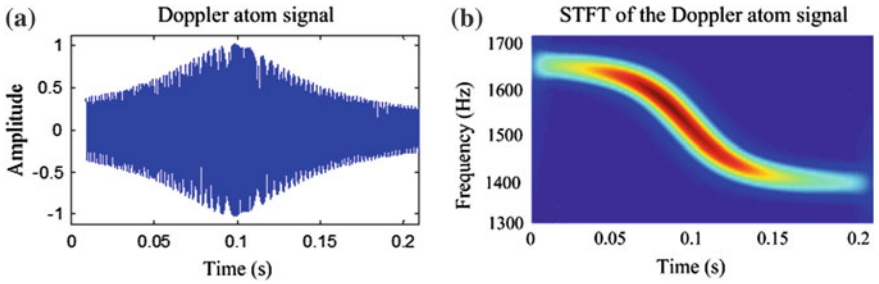


Fig. 2 **a** waveform of a Doppler atom signal; **b** STFT of the Doppler atom signal

$$P_{v_0,r,S,c,f_o,\varphi} = \frac{2\pi q_0 f_0}{4\pi R(1 - M\cos\theta)^2}\cos(2\pi q_0 f_0 t + \varphi) \tag{3}$$

which we call the "Doppler atom". Figure 2a depicts a waveform of such a Doppler atom signal, with its STFT spectrum in Fig. 2b. Clearly, such a Doppler atom is dictated by the following six parameters

$$\gamma = (v_0, r, S, c, f_o, \varphi) \tag{4}$$

3 Introduction of Correlation Filtering Analysis

Recently, Wang et al. [10] used the wavelet correlation filtering to find a suitable parametric wavelet model to match the bearing fault transient. Then, they constructed a strict periodic multiple-transient model to detect the temporal cyclic intervals of bearing localized faults.

The idea of correlation filtering analysis is to calculate the Pearson's correlation coefficient between the collected Doppler-shifted signal and the parametric atom. The Pearson's correlation coefficient is a powerful tool to measure the strength of the linear dependence between two signals. Denote the collected Doppler-shifted signal as $x(n)$ and the parametric atom as $P_\gamma(n)$. Assume both of the signals has a zero mean. The Pearson's correlation coefficient $CF_{x(n),P\gamma(n)}$ between them can be defined as: Stigler [11]

$$CF_{x(n),P_\gamma(n)} = \frac{< x(n), P_\gamma(n) >}{\sqrt{x(n), x(n)}\sqrt{P_\gamma(n), P_\gamma(n)}} = \frac{\sum_{i=1}^{N} x(i) \cdot P_\gamma(i)}{\sqrt{\sum_{i=1}^{N} x(i) \cdot x(i)}\sqrt{\sum_{i=1}^{N} P_\gamma(i) \cdot P_\gamma(i)}} \tag{5}$$

where N is the length of the signal and $<\cdot>$ is the inner product. In terms of Cauchy-Schwarz inequality, the Pearson's correlation coefficient is constrained to:

$$-1 \leq CF_{x(n),P_{\hat{\gamma}}(n)} \leq 1 \tag{6}$$

The closer the correlation coefficient is to 0, the weaker the linear dependence relationship between the two signals.

4 Introduction of Doppler Effect Reduction Based on Re-sampling Method

As shown in Fig. 1, the sound source is a continuous analog signal while the observer is a microphone which is static to the air medium. The sound source moves from point A to B, with a constant speed of V_0. Point A represents the initial position where the time t is set to be zero.

The amplitude weight radiated from the source at time t_r (emit time) will arrive at the microphone at time t_R (receive time)

$$t_R = t_r + \sqrt{r^2 + [S/2 - v_0 t_r]^2}/c \tag{7}$$

which can be used as the time vector for interpolation. After interpolating the recorded Doppler-shifted signal with t_R, an amplitude vector x_r is obtained, and finally the restored signal is $x_r(t_r)$.

5 An Adaptive Doppler Effect Reduction Algorithm for the ADBD System

Generally, steps of the proposed adaptive Doppler effect reduction algorithm are as follows:

(1) Calculate the FFT spectrum of the recorded Doppler-shifted signal x_{dop}, and select the sensitive frequency band by inspecting the FFT spectrum and discretize it to $\{f_c(i), i = 1,2,\ldots,N\}$;
(2) Initialize $i = 1$. Apply the band-pass filter on x_{dop} at the center frequency of $f_c(i)$ to get a narrow-band Doppler shift signal $x_{nar}(i)$;
(3) Search the parameters space to find the Doppler atom which matches $x_{nar}(i)$ best by the correlation filtering analysis and obtain the optimal parameters $\gamma(i)$ corresponding to the maximum Pearson's correlation coefficient $CF(i)$;

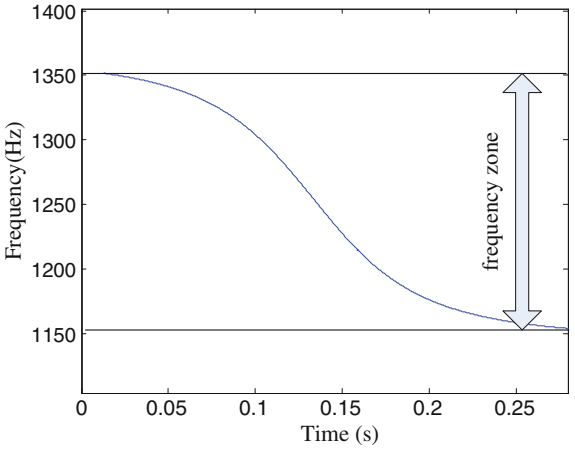


Fig. 3 Curve of frequency variation of a Doppler-shifted signal with center frequency of 1245 Hz

(4) Let $i = i+1$, and repeat Step(2) ~ (3) until $i = N$, then we get $\{\gamma(i), i = 1,\ldots,N\}$ and $\{CF(i), i = 1,\ldots,N\}$. The parameters set γ^{opt} corresponding to the maximum $CF(i)$ is chosen as the optimal parameters set;
(5) Restore the Doppler-shifted signal $\{x_{dop}\}$ by the re-sampling method using the optimal parameters γ^{opt}

It should be noted that zone of the band-pass filter in Step (2) will highly influence the accuracy of the reduction result. Considering a Doppler-shifted signal with a single characteristic frequency f_0, shown as Eq. (3), the variation of the frequency could be attained by the following equation:

$$f = f_0 \frac{M(S - vt) + \sqrt{(S - vt)^2 + (1 - M^2)r^2}}{(1 - M^2) \cdot \sqrt{(S - vt)^2 + (1 - M^2)r^2}} \tag{8}$$

Figure 3 shows the instantaneous frequency variation of the Doppler-shifted signal with a center frequency $f_0 = 1245$ Hz.

Then, the frequency zone of the band-pass filter can be determined by the following steps:

(1) The center frequency with the maximum power is selected and be written as f_0 by inspecting the frequency spectrum of the original Doppler-shifted bearing fault signal;
(2) Calculate the frequency variation f by Eq. (8), then the frequency zone of the band-pass filter is $[min(f)\ max(f)]$.

According to Eq. (4), there are totally six parameters, so the computation is quite expansive. Ignoring the influence of temperature and air pressure, the propagation speed of sound in air is c = 340 m/s, thus the parameters estimation space demotes to five dimensions $\gamma = (v_0, r, S, f_0, \varphi)$.

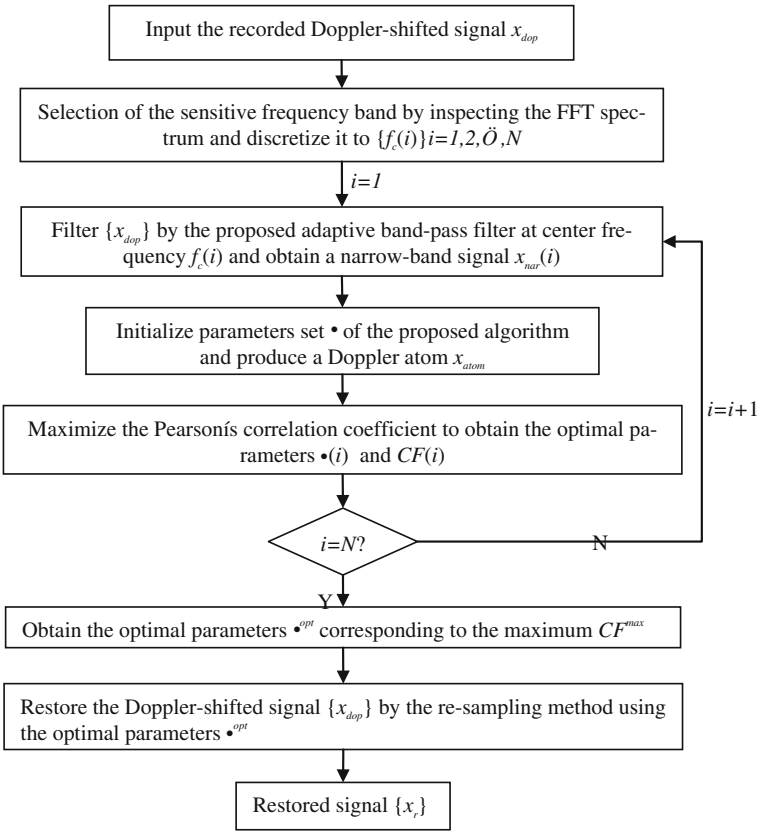


Fig. 4 Flowchart of adaptive Doppler effect reduction algorithm

The flowchart of the adaptive Doppler effect reduction algorithm is shown in Fig. 4.

6 Experimental Verification

To verify the effectiveness of the proposed adaptive Doppler effect reduction algorithm for the wayside fault diagnosis of train bearings acoustic signals, an experiment was implemented to obtain the Doppler-shifted signal of the bearing (Type: NJ(P)3226X1), which is the dominated type in use. Some of the parameters of the bearing is shown in Tables 1 and 2.

The artificial cracks had been set by the wire-electrode cutting machine with the width of 0.18 mm on the outer race. Two steps had been implemented to get the Doppler-shifted signal of the defective bearing. First, the defective bearing is tested

Table 1 Specification of the testing bearing

Type	Diameter of the outer race	Diameter of the inner race	Pitch diameter (D)	Diameter of the roller (d)	Number of the roller (z)
NJ(P) 3226XI	250 mm	130 mm	190 mm	32 mm	14

Table 2 Parameters in the experiment

Load	Rotating speed	Sampling frequency
3*t*	1450 rpm	50 kHz

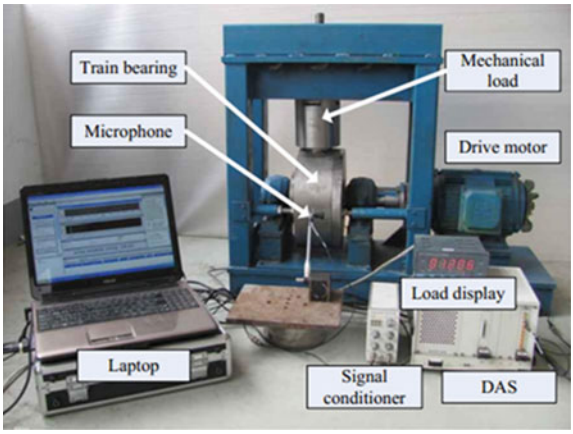


Fig. 5 Experimental bench

by the bench shown in Fig. 5 under the radial load of 3*t*. The rotation speed of the motor is set to be 1,430 rpm, and the sampling frequency is 50 kHz. The acoustic signal was acquired by a microphone (Type: 4944-A, Manufactory: B&K Company) mounted beside the outer race of the bearing. The acoustic signal acquired through the microphone was preprocessed by the signal conditioner and then recorded by the data acquisition devise (DAS). Some of the parameters during the experiment is shown in Table 1.

The recorded signal (Doppler free) was then played by a loudspeaker fixed on a moving vehicle. As a result, the signal acquired according to the microphone sited beside the motion trail of the vehicle is the original signal with Doppler effect. The experiment can be referred to Fig. 1 with the following geometric parameters: $S = 4$ m, $r = 2$ m, $V_0 = 30$ m/s.

The waveform of the recorded Doppler-shifted signal emit from the testing bearing with a single defect on the outer race is shown in Fig. 6a with its envelope spectrum in Fig. 6b. Then the proposed adaptive Doppler effect reduction algorithm is applied to analyse the Doppler-shifted signal.

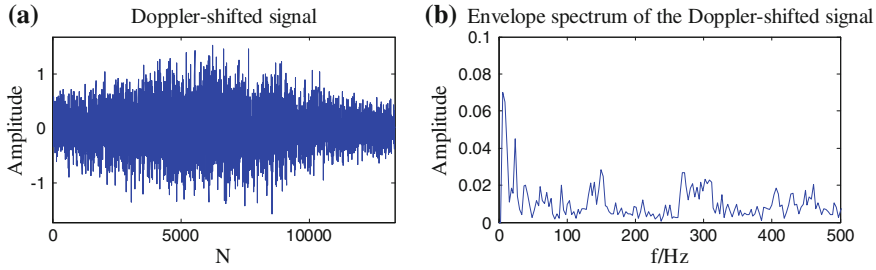


Fig. 6 **a** Waveform of the Doppler-shifted signal emit from the testing bearing with a single defect on the outer race; **b** Envelope spectrum of the Doppler-shifted signal

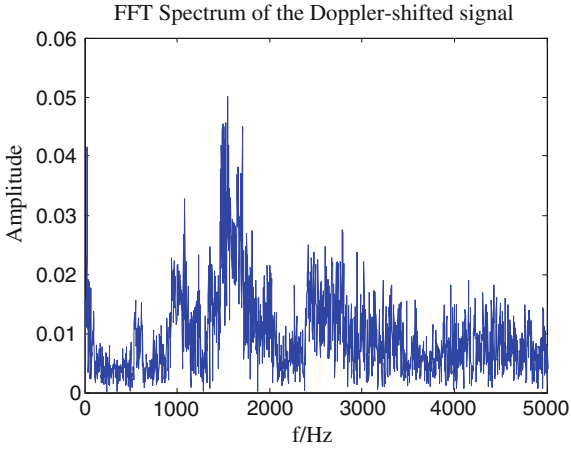


Fig. 7 FFT spectrum of the Doppler-shifted signal emit from the testing bearing with a single defect on the outer race

First, by inspecting the FFT spectrum in Fig. 7, the frequency band of [1200 Hz 1800 Hz] is selected as the sensitive frequency band, and then this frequency band is discretized to 600 point, which means the resolution of the searching result for the center frequency is 1 Hz.

Then the source Doppler-shifted signal is firstly filtered by a Butterworth band-pass filter. Then searching the parameters space to find the Doppler atom which matches the filtered signal best by the correlation filtering analysis. After all frequency values in the sensitive frequency band have been analysed, the optimal parameters corresponding to the maximum Pearson's correlation coefficient was obtained. The optimal parameters are shown in Table 3.

Finally, the time vector for re-sampling (shown in Fig. 8a) is calculated according to the optimal parameters. And the Doppler-shifted signal is restored using the re-sampling method. The waveform of the restored signal is shown in Fig. 8b. The FFT spectrum and the envelope spectrum of the restored signal are shown in Fig. 9a and b respectively. In comparison with the original signal's FFT spectrum (Fig. 7), the FFT spectrum of the restored signal is much more sharp

Table 3 Optimal parameters γ^{opt} corresponding to the maximum CF^{max}

V_0	r	S	f_0	φ
31.5 m/s	1.9 m	4.2 m	1572 Hz	0.48 rad

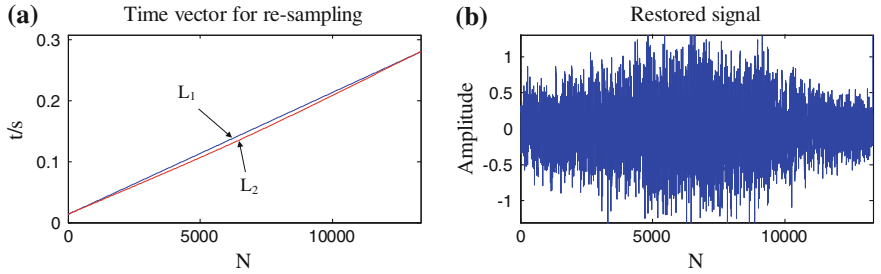


Fig. 8 **a** L1-Time vector of the original Doppler-shifted signal emitted from the testing bearing with a single defect on the outer race; L2-Time vector for re-sampling; **b** Waveform of the restored signal

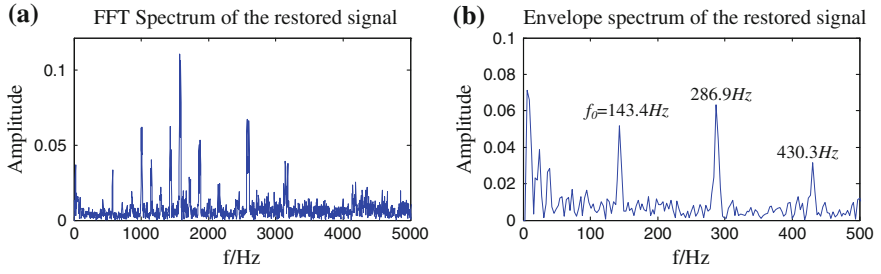


Fig. 9 **a** FFT spectrum of the restored signal; **b** Envelope spectrum of the restored signal

which means that the frequential structure disturbance resulting from the Doppler effect is removed clearly. And it is obvious to see in the envelope spectrum of the restored signal (Fig. 9b) that the failure frequency and its second and third harmonic are very sharp, narrow and accurate, which means that the signal is restored perfectly.

7 Conclusion

An adaptive Doppler effect reduction algorithm for the wayside ADBD system is proposed to remove the Doppler effect embedded in the recorded acoustic bearing signal. The parameters of the Doppler kinematic model are estimated by correlation filtering analysis and the Doppler effect is removed then by the re-sampling method.

The effectiveness of the method is verified by an experimental case using a train bearing with a single defect on the outer race, and the result shows that the frequential structure of the Doppler-shifted signal has been corrected perfectly, the phenomenons such as frequency band expansion and frequency shift have been removed clearly which allows much more efficient diagnosis for the train bearing.

Compared with the methods based on IFE, this method does not need to know the characteristic frequency beforehand. Compared with the time domain methods this method does not need to know the geometric parameters beforehand. So it is greatly beneficial to performance enhancement of the ADBD system.

Acknowledgment This work is supported by the National Natural Science Foundation of China (No. 51075379, No. 51005221) and partly by the Natural Science Major Project of Education Department of Anhui Province (No. KJ2013A010).

References

1. Choe HC, Wan YL, Chan AK (1997) Neural pattern identification of railroad wheel-bearing faults from audible acoustic signals: comparison of FFT, CWT and DWT features. SPIE Proc Wavelet Appl 3087:480–496
2. Sneed WH, Smith RL (1998) On-board real-time bearing defects detection and monitoring. Proceedings of the 1998 ASME/IEEE joint railroad conference, 1998. pp 149–153
3. Irani FD et al (2002) Development and deployment of advanced wayside condition monitoring systems. Foreign Rolling Stock 39(2)
4. Stojanovic M, Catipovic JA, Proakis JG (1994) Phase-coherent digital communications for underwater acoustic channels. IEEE J Oceanic Eng 19:100–111
5. Johnson M, Freitag L, Stojanovic M (1997) Improved doppler tracking and correction for underwater acoustic communications. IEEE Int Conf Acoust, Speech, Signal Proc 1:575–578
6. Dybała J, Radkowski S (2013) Reduction of doppler effect for the needs of wayside condition monitoring system of railway vehicles. Mech Syst Signal Process 38(1):125–136
7. Yang DG, Wang ZT (2011) Quantitative measurement of pass-by noise radiated by vehicles running at high speeds. J Sound Vib 330:1352–1364
8. Morse PM, Ingard KU (1987) Theoretical acoustics. Princeton
9. Yang DG, Zheng SF, Luo YG, Lian XM, Jiang XY (2002) Acoustic holography method for the identification of moving sound source. ACTA Acustica 27(4):257–362
10. Wang S, Huang W, Zhu ZK (2011) Transient modeling and parameter identification based on wavelet and correlation filtering for rotating machine fault diagnosis. Mech Syst Signal Process 25:1299–1320
11. Stigler SM (1989) Francis Galton's account of the invention of correlation. Stat Sci 4:73–79
12. Cline JE, Bilodeau JR (1998) Acoustic wayside identification of freight car roller bearing detects. Proceedings of the 1998 ASME /IEEE joint railroad conference, pp 79–83
13. Barke D, Chiu WK (2005) Structural health monitoring in the railway industry: a review. Struct Health Monit 4:81–93

Lamb Waves Inspection by Using Chirp Signal and Mode Purification

Zenghua Liu, Yingzan Xu, Cunfu He and Bin Wu

Abstract Ultrasonic Lamb wave inspection is one of nondestructive testing approaches, and it has many advantages including large-area, omnidirectional, fast inspection. In common, narrow-band frequency sinusoidal tone burst signal modulated by window function is used to excite guided waves in waveguides for mode control and tuning. For obtaining the signals at different frequencies, guided wave inspection will need to be repeated many times. As an alternative of narrow-band guided waves excitation, a broadband chirp signal is used for excitation signal and the received signal is post-processed to obtain any one single frequency tone burst signal which frequency range is located in the frequency bandwidth. This approach greatly increases guided waves inspection efficiency and ensures the consistencies. As the receivers, two PZT elements are supposed to be attached on both sides of a PMMA plate in the same location. The received signals from two PZT elements are summed to obtain single S_0 mode, and are subtracted to obtain single A_0 mode achieving mode purification of Lamb waves. Ultrasonic Lamb wave technology is applied to inspect the plate-like structures using chirp signal and mode purification. This method is more efficient and accurate to obtain data and much easier to identify, interpret and extract the information of the defects.

Keywords Lamb waves · Chirp signal · Mode purification · Defect inspection · Frequency band

1 Introduction

Lamb waves are specific type of elastic stress waves travelling in the plate-like structures which contributes both longitudinal and shear partial wave components. On the basis of vibration modes of particles in the plate, Lamb waves have two

Z. Liu (✉) · Y. Xu · C. He · B. Wu
College of Mechanical Engineering and Applied Electronics Technology,
Beijing University of Technology, Beijing, China
e-mail: liuzenghua@bjut.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_14

types of modes: symmetric (S) mode and antisymmetric (A) mode. Because of the advantages of inspection of large areas, little energy attenuation and excellent sensitivity to multiple defects, Lamb waves are suitable for defects detection in the plate-like structures [1, 2] and it is widely applied in the nondestructive testing (NDT) and structural health monitoring (SHM) [3–6].

The traditional ultrasonic Lamb waves exciting signal is sinusoidal tone burst signal modulated by window function which has narrow frequency bandwidth. Duration time of this type of excitation signal is very short and signal energy is relatively concentrated in narrow frequency range so that guided wave dispersion can be effectively restrained and therefore the sensitivity to sorts of defects can be improved [7, 8]. On the occasions where a large amount of data acquisition is required, it is much time consuming and is high demanding for the data acquisition instruments using narrow-band tone burst signal. Besides, the long duration of measurement increases the inconsistencies of measurement conditions and eventually affects defect identification. To address this problem, with a chirp signal being transmitted, the received signal is post-processed to obtain any one narrow-band tone burst signal which frequency range is located in the frequency bandwidth. This approach greatly increases inspection efficiency and ensures the consistencies [9]. The more pure Lamb wave mode is, the easier defect identification is. Therefore, many researchers apply themselves to find a method that can get single Lamb waves mode. For instance, Clarke et al. [10, 11] studied the effect of dimension parameters of PZT elements on the resonance frequency and frequency bandwidth and eventually established the optimal transducer geometry that can generate a high pure A_0 mode at low frequencies. Su and Ye [12] attached two identical PZT elements on both sides of the aluminium plate in the same location. These elements were applied for in-phase and out-of-phase signal simultaneously so that single S_0 or A_0 mode can be excited, respectively.

In this paper, broadband chirp is used as excitation signal. The received signal is post-processed to obtain any one narrow-band tone burst signal which frequency range is located in the frequency bandwidth. Two identical PZT elements are attached on both sides of the plate in the same location. The signals received from two PZT elements are summed to obtain single S_0 mode, and are subtracted to obtain single A_0 mode achieving mode purification. Ultrasonic Lamb wave technology is applied to inspect the plate-like structures using chirp signal and mode purification.

2 Chirp Excitation of Lamb Waves

Chirp is a kind of linear frequency modulated signal with a certain frequency bandwidth and duration time. Its frequency linearly increases from the minimum value to the maximum value at a certain step length while remaining the amplitude unchanged. The mathematical expression is

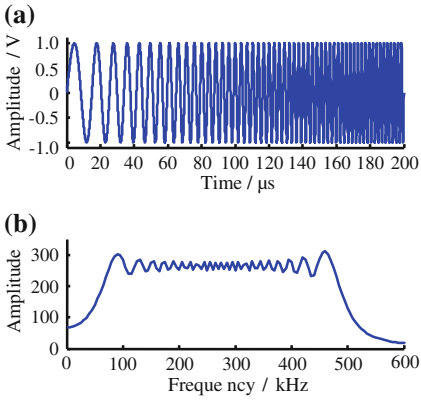


Fig. 1 Chirp excitation signal, **a** Time domain waveform, **b** Spectrum diagram

$$s_c(t) = w(t)\sin\left(2\pi f_0 t + \frac{\pi Bt^2}{T}\right) \tag{1}$$

where $w(t)$ is a rectangular-window function, f_0 is the starting frequency, B is the frequency bandwidth, and T is the chirp duration time. Figure 1 illustrates a chirp excitation signal which the frequency sweeps from 30 to 500 kHz and the duration is over 200 μs. For the comparison, Fig. 2 shows a tone burst excitation signal modulated by Hanning window at 200 kHz and with a duration of 5 cycles.

The general process of Lamb wave inspection is as follows [9, 13]: a transmitter excites Lamb waves travelling in the plate. It assumes that there are some defects in the plate, Lamb waves are supposed to scatter encountering the defects. Then, the scattered Lamb wave signals are received by a receiver, eventually the received signals are collected by data acquisition instruments. The entire detection system consists of excitation signal source, transmitters, tested plates, receivers, data acquisition instruments and computer. This system can be modelled as a linear

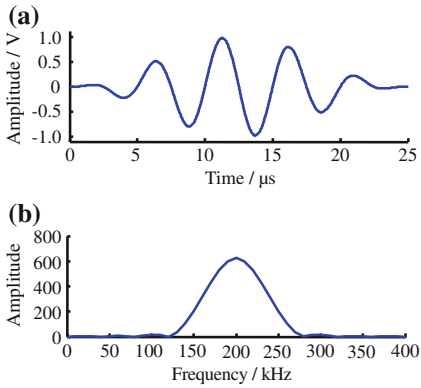


Fig. 2 Tone burst excitation signal, **a** Time domain waveform, **b** Spectrum diagram

system. Therefore, the response to the chirp excitation can be expressed in the frequency domain as

$$R_c(\omega) = H(\omega)S_c(\omega) \tag{2}$$

where $R_c(\omega)$ is the Fourier transform of a chirp receiving signal, $S_c(\omega)$ is the Fourier transform of a chirp excitation signal, and $H(\omega)$ is the frequency response function of detection system.

Similarly, the response to a narrow-band tone burst excitation signal is supposed to be showed as Eq. 3. Here, $S_c(\omega)$ is different with $S_t(\omega)$ because of only the different excitation signal. However, $H(\omega)$ is consistent due to the same detection system.

$$R_t(\omega) = H(\omega)S_t(\omega) \tag{3}$$

Through H, we can connect Eqs. 2 and 3 as follows

$$H(\omega) = \frac{R_c(\omega)}{S_c(\omega)} = \frac{R_t(\omega)}{S_t(\omega)} \tag{4}$$

Eventually, R_t can be obtain expressed as

$$R_t(\omega) = S_t(\omega)\frac{R_c(\omega)}{S_c(\omega)} \tag{5}$$

which is the Fourier transform of the extracted tone burst signal form chirp receiving signal. Its inverse Fourier transform is the time domain signal we intend to obtain and it basically equals to the direct response of narrow-band tone burst excitation signal.

The great advantage of chirp excitation signal is that we can get any one narrow-band tone burst signal which frequency range is located in the frequency bandwidth with a chirp being excited in Lamb wave inspection. The extracted signal almost equals to the response of narrow-band tone burst signal with same center frequency. Thus, it is easy to achieve received signals with arbitrary center frequency by triggering a single excitation signal and therefore greatly improves the detection efficiency.

3 Mode Purification Theory of Lamb Waves

An important issue to be solved is mode selection and mode purification in Lamb wave inspection technology. Because the sensitivities of different modes are different to sorts of defects, and multiple modes would make the received signals much more complicated so that sometimes it is difficult to interpret [14]. To address

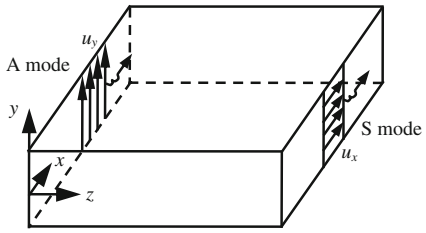


Fig. 3 Motion forms of Lamb wave modes

this problem, this section describes a method that can achieve mode purification of Lamb waves.

The vibration forms of two Lamb wave modes are different. In Fig. 3, it is clear to see that both S mode and A mode propagate in the x direction. The displacement of S mode is mainly u_x, in the x direction and meanwhile the displacement of A mode concentrates on the y direction, u_y.

To understand the behaviours of S_0 mode and A_0 mode better, their displacement distributions of these fundamental modes in a 4.8 mm-thick PMMA plate are calculated by Disperse software shown in Fig. 4a and b, respectively. According to the wave structures shown in these figures, it can be seen that displacement components of S_0 mode and A_0 mode consist of in-plane displacement and out-of-plane displacement. It can be also noticed that S_0 mode has mainly in-plane displacement and A_0 mode has mainly out-of-plane displacement. Furthermore, as is known that particle motion of S_0 mode is symmetric regard to the middle plane of the plate. However, it is opposite for A_0 mode.

According to the characteristics of wave structure of Lamb waves, two PZT elements where are located on both surfaces of the plate in the same location have in-plane displacement and out-of-plane displacement. Their in-plane displacements are identical in both amplitude and phase. However, the out-of-displacements own the same amplitude and inverse phase. Therefore, the received signals from two PZT elements are summed to obtain single S_0 mode, and are subtracted to obtain single A_0 mode achieving mode purification of Lamb waves. The amplitudes of single Lamb waves can be described as follows

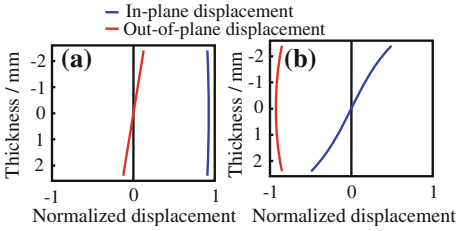


Fig. 4 Displacement distributions of Lamb waves at 50 kHz, **a** S_0 mode, **b** A_0 mode

$$r_S = \frac{r_{\text{upper}} + r_{\text{lower}}}{2} \tag{6}$$

$$r_A = \frac{r_{\text{upper}} - r_{\text{lower}}}{2} \tag{7}$$

where r_S is the amplitude of purified S_0 mode, r_A is the amplitude of purified A_0 mode, r_{upper} and r_{lower} are two received signals from two PZT elements where are located on both surfaces of the plate in the same location.

Mode purification of Lamb waves is very beneficial to defect identification in the plate-like structures. The transmitters generally excite multiple modes simultaneously including fundamental S and A modes even higher orders. But the received signals can be post-processed to achieve mode purification of Lamb waves so that it can be much easier to identify, interpret and extract the information of the defects.

4 Experimental Research

With a broadband chirp signal being excited in the experiment, it is feasible to extract any one narrow-band tone burst signal which frequency range is located in the frequency bandwidth from the received signal. The extracted narrow-band tone burst signals are purified to obtain single Lamb wave mode thus detecting defects more efficiently and easily.

5 Excitation and Extraction of Chirp Signal

Experiment testing system consists of an arbitrary waveform generator (AFG3021B), PZT elements, a PMMA plate, a digital oscilloscope (DPO4054) and a computer, the schematic diagram is shown in Fig. 5. The exciting PZT element and receiving PZT element are 290 mm apart away which diameters are 12 mm and thicknesses are 1.5 mm. The geometric dimensions of PMMA plate are that the length is 1 m, the width is 1 m and the thickness is 4.8 mm.

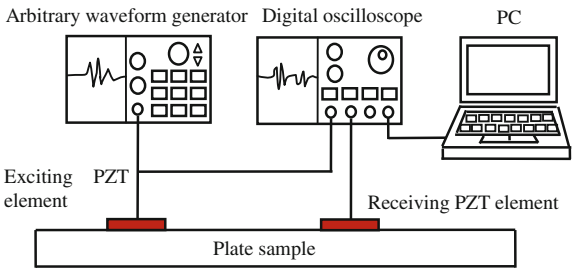


Fig. 5 Schematic diagram of experiment testing system using a chirp excitation signal

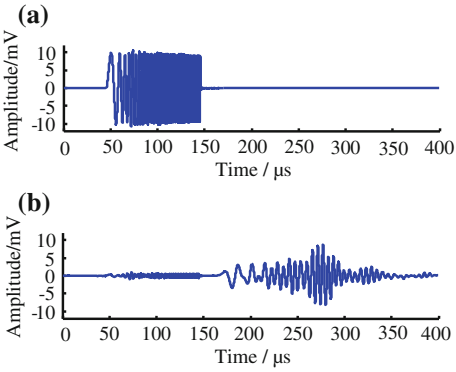


Fig. 6 Chirp excitation signal and response signal, **a** Chirp excitation signal, **b** Response signal

In the experiment, a broadband chirp signal is selected to be the excitation signal which frequency bandwidth is 30–900 kHz, duration time is 100 μs and peak-to-peak amplitude is 10 V. The chirp signal triggered by an arbitrary generator is applied to an exciting PZT element to produce Lamb waves travelling in the plate. Lamb waves are received by a receiving PZT element, eventually received signals are obtained by the digital oscilloscope. The chirp excitation signal and response signal are shown in Fig. 6a and b, respectively.

It can be seen from Fig. 6b that the chirp response signal is indecipherable. Because the chirp is a linear frequency-modulated signal and the response signal has to be processed to extract narrow tone burst signals. Using the method in Sect. 2, narrow-band tone burst signals with 5 cycles are extracted from this chirp response signal. These tone burst signals are centered at 50, 75, 100, 125, 150, 175 and 200 kHz, respectively. These extracted signals are compared with those obtained by using narrow-band tone burst signals as excitation signals. For the verification of effectiveness of this extraction method, Fig. 7 gives the comparison of direct received signals by using tone burst excitation signals and chirp extracted signals.

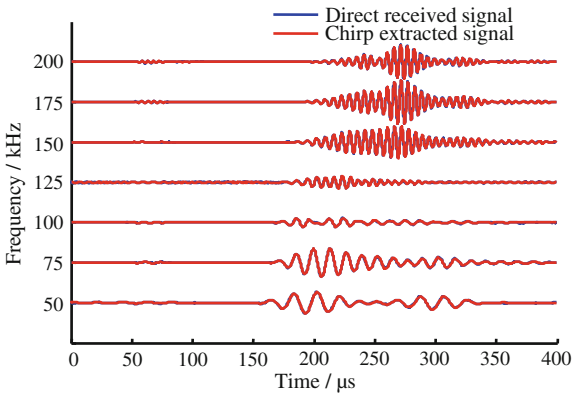


Fig. 7 Comparisons of direct received signals of tone burst signals and chirp extracted signals at different frequencies

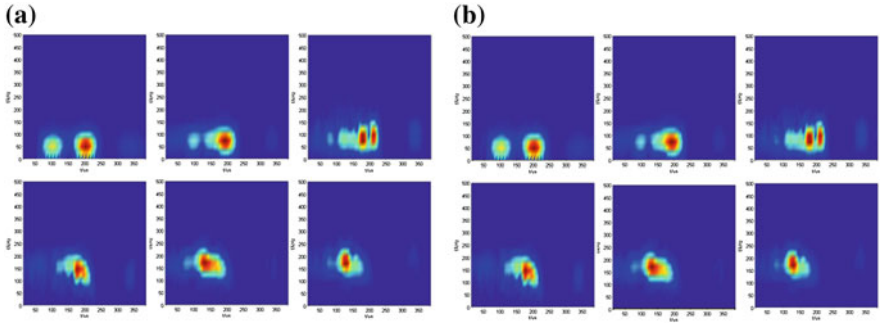


Fig. 8 Time-frequency spectra of direct received signals and chirp extracted signals, **a** Direct received signals, **b** Chirp extracted signals

As is clearly seen from Fig. 7 that the chirp extracted signals match very well with direct received signals of tone burst signals suggesting that the broadband chirp signal can accurately extract narrow-band signals which frequency range is located in the frequency bandwidth of chirp signal. The response signal at 50 kHz has two Lamb wave modes away clearly with low dispersion and high concentrated energy. These factors are greatly beneficial to detect defects and improve the detection accuracy in the plate-like structures.

The time-frequency analysis is applied to both the chirp extracted signals and PC direct received signals of tone burst signals to verify the consistency of frequency components and energy distribution for them. Figure 8a and b are the time-frequency spectra of direct received signals and chirp extracted signals, respectively. The center frequencies of these direct received signals and chirp extracted signals are 50, 75, 100, 125, 150, 175, and 200 kHz, respectively. It is clear from Fig. 8 that two kinds of signals basically have the same frequency components and energy distribution at the same frequency. Therefore, with a chirp signal being transmitted, the received signal is post- processed to obtain any one narrow-band tone burst signal which frequency range is located in the frequency bandwidth.

At present, a chirp response signal is processed to get narrow-band tone burst signals at 50 kHz ranging from 3 to 9 cycles. Similar as above, the comparison of chirp extracted signals with direct received signals of tone burst excitation signals is shown in Fig. 9. The perfect match suggests that multiple narrow-band tone burst signals with different cycles can be extracted from a chirp response signal. In terms of interpreting these response signals, it is clearly seen that S_0 mode and A_0 mode are trended to overlap and endure more serious dispersion along with the increasing numbers of cycles. These factors are disadvantageous for detecting defects in the plates.

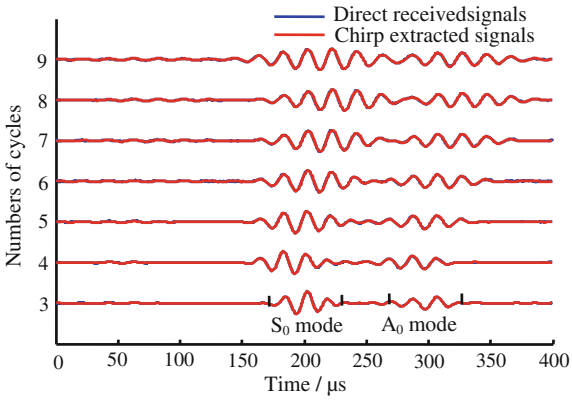


Fig. 9 Comparisons of direct received signals of tone burst signals and chirp extracted signals with different numbers of cycles

6 Mode Purification of Lamb Waves

Experiment testing system is similar to that above there being only one distinction where Lamb waves are received by two identical receiving PZT elements attached on both sides of the plate in the same location. Figure 10 illustrates explicitly the schematic diagram of experiment testing system applying in this section.

Figure 11 shows two extracted narrow-band tone burst signals from PZT elements attached on both surfaces in the plate. These signals are Hanning window-modulated sinusoidal signals with 5 cycles at 50 kHz. It can be confirmed that the first wave package is S_0 mode and the second is A_0 mode. What are worth noticing is that the S_0 modes of two signals are very same each other, however A_0 mode has identical amplitudes but inverse phases. This result is well in accordance with propagation characteristics of Lamb waves. According to these characteristics, the extracted signals are post-processed to achieve mode purification of Lamb waves shown in Fig. 12. There is no doubt that single S_0 mode and single A_0 mode could be separated and purified easily using the mode purification method introduced above.

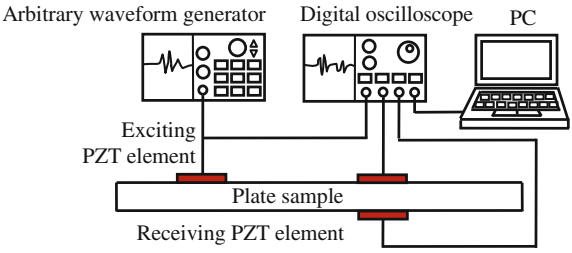


Fig. 10 Schematic diagram of experiment testing system for mode purification of Lamb waves

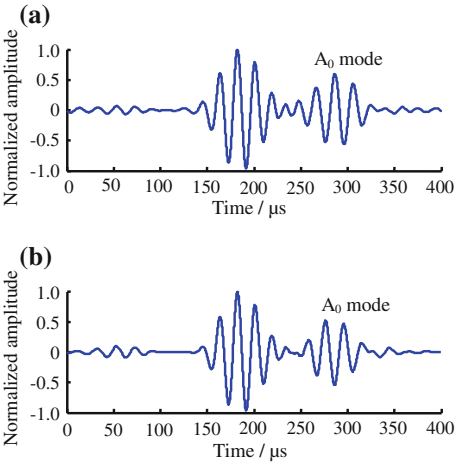


Fig. 11 Two chirp extracted signals from PZT elements attached on both surfaces in the plate, **a** On the *top* surface, **b** On the *lower* surface

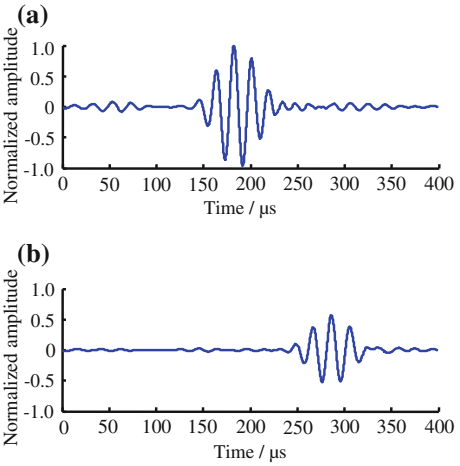


Fig. 12 Mode purification of Lamb waves, **a** Purified S_0 mode, **b** Purified A_0 mode

A same chirp signal as above is applied to the experiment testing system as the excitation signal. Lamb waves generated by arbitrary waveform generator travel in the plate and eventually are received by two PZT elements attached on both sides of the plate in the same location. Firstly, narrow-band tone burst signals are extracted from two chirp received signals. Then narrow-band tone burst signals are post-processed to achieve mode purification of Lamb waves.

The existence of multiple modes makes received signals more complicated so that it is more difficult to identify, interpret and extract the characteristic signals in Lamb waves testing technology. Mode purification method of Lamb waves exactly can solve this problem and is very potential to apply in NDT and SHM.

7 Conclusions

Utilizing a broadband chirp excitation signal in Lamb wave inspection, the received signals are post-processed to obtain narrow-band tone burst signals which frequency band are in the frequency bandwidth and duration times do not exceed the duration time of chirp signal. This method greatly increases detection efficiency and ensures the consistencies of detection conditions.

On the basis of the propagation characteristics of Lamb waves, this work proposes the mode purification method of Lamb waves achieving modes separation and purification. With the purified single mode, it is supposed to be much easier to defect identification.

Acknowledgments This work was supported by the National Natural Science Foundation of China (Grant Nos. 11272021 and 50975006), Beijing Natural Science Foundation (Grant No. 1122007), the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions (No. CIT&CD201304048) and Beijing Nova Program (Grant No. 2008A015).

References

1. Michaels JE, Michaels TE (2007) Guided wave signal processing and image fusion for in situ damage localization in plates. Wave Motion 44:482–492
2. Zhao X, Roger L, Owens S et al (2011) Ultrasonic lamb wave tomography in structural health monitoring. Smart Mater Struct 20:105002(10)
3. Liu Z, Yu H, He C et al (2013) Delamination damage detection of laminated composite beams using air-couple ultrasonic transducers. Sci China: Phys, Mech Astron 56:1269–1279
4. Giurgiutiu V (2005) Tuned Lamb wave excitation and detection with piezoelectric wafer active sensors for structural health monitoring. J Intell Mater Syst Struct 16:291–305
5. Michaels JE (2008) Detection, localization and characterization of damage in plates with an in situ array of spatially distributed ultrasonic sensors. Smart Mater Struct 17:035035(10)
6. Liu Z, Yu F, Wei R et al (2013) Image fusion based on single-frequency guided wave mode signals for structural health monitoring in composite plates. Mater Eval 71:1434–1443
7. Wilcox P, Lowe M, Cawley P (2001) The effect of dispersion on long-range inspection using ultrasonic guided waves. NDT & E Int 34:1–9
8. Wilcox P, Dalton R, Lowe M (2001) Mode and transducer selection for long range Lamb wave inspection. J Intell Mater Syst Struct 12:553–565
9. Michaels JE, Lee S, Croxford A et al (2012) Chirp excitation of ultrasonic guided waves. Ultrasonics 53:265–270
10. Clarke T, Simonetti F, Rohklin S et al (2009) Development of a low-frequency high purity A_0 mode transducer for SHM applications. IEEE Trans Ultrason Ferroelectr Freq Control 56:1457–1468
11. Clarke T, Cawley P (2010) Enhancing the defect localization capability of a guided wave SHM system applied to a complex structure. Struct Health Monit 10:247–259
12. Su Z, Ye L (2004) Selective generation of Lamb wave modes and their propagation characteristics in defective composite laminates. J Mater: Des Appl 218:95–110
13. Zheng Y, He C, Wu B (2013) Chirp signal and its application in ultrasonic guided wave inspection. Chin J Sci Instrum 34:552–558 (in Chinese)
14. Alleyne D, Cawley P (1992) Optimisation of Lamb wave inspection techniques. NDT & E Int 25:11–22

Performance Degradation Assessment of Slurry Pumps

Peter W. Tse and Dong Wang

Abstract Slurry pumps are widely used in oil sand pumping operations to enhance the potential and kinetic energy of liquid and solid mixtures and pump the mixtures from one place to another place. The rotating impellers of slurry pumps operate continuously and they are unavoidably abraded and eroded by the transferring liquids and solids. Therefore, impeller wear is one of the major causes for slurry pump breakdown. In order to ensure the high reliability of the use of impellers and prevent the occurrence of impeller failures, the performance degradation assessment of impellers is necessary to be investigated. In this paper, a moving-average mean wear degradation index and a moving-average deviation wear degradation index are proposed to track the health condition of the impellers used in oil sand pumps. The influence of different parameters on the performance degradation assessment is discussed. The vibration signals collected from an industrial oil sand pump are used to validate the proposed impeller heath indicators. The results show that the proposed health indicators are effective in describing the impeller health state evolution.

1 Introduction

Slurry pumps are common machines used in oil sand pumping operations to enhance the potential and kinetic energy of liquid and solid mixtures. They aim to pump the mixtures from one place to another place. The rotating impellers of slurry pumps operate continuously and they are unavoidably abraded and eroded by the transferring liquids and solids. As a result, severe impeller wear is one of the major causes for slurry pump breakdown and thus impeller health prognosis must be investigated. The health prognosis of impellers means to estimate the remaining

P.W. Tse (✉) · D. Wang
Department of Systems Engineering and Engineering Management,
City University of Hong Kong, Kowloon Tong, Hong Kong, China
e-mail: meptse@cityu.edu.hk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_15

useful life of impellers. Prior to the estimation of remaining useful life, it is necessary to construct a health indicator used for describing the degradation trend of impellers. In other words, impeller performance degradation assessment is a basis for impeller remaining useful life estimation. In recent years, performance degradation assessment attracts much attention. Qiu et al. [1] used optimal Morlet wavelet for bearing fault feature extraction and employed self organizing map for bearing performance degradation assessment. Wang et al. [2] used discrete wavelet transform for gear performance degradation assessment. Ocak et al. [3] used the decrease of the probabilities of hidden Markov model trained by the wavelet packet node energy features extracted from normal bearing data to describe the bearing degradation. The similar idea was also applied by Miao et al. [4] to track the gear health condition. Prior to the use of hidden Markov model, empirical mode decomposition was employed to extract gear fault features from normal gear data. Hong and Liang [5] used Lempel–Ziv complexity to assess bearing performance degradation. Pan et al. [6] used wavelet packet node energies as bearing fault features to train support vector data description that was used to track bearing health condition. Wang et al. [7] used a series of wavelet filters to extract gear fault features and applied support vector data description to assess gear health condition. The variations of support vector data description, such as fuzzy support vector data description [8] and rough support vector data description [9], were also used for bearing performance degradation assessment. Pan et al. [10] also employed the combination of fuzzy c-means and support vector data description to track bearing health condition. Yu [11] used locality preserving projections to describe bearing performance degradation. Miao et al. [12] used multiple wavelet filters to extract low frequency fault features for fan bearing performance degradation assessment. The literatures on impeller performance degradation assessment are few. In recent years, Wang et al. [13] artificially introduced some impeller damage modes and their different wear degrees to normal impellers. Then, vibration data was collected from experiment systems combined with these fault impellers. A joint method of support vector machine, a novel data cleaning algorithm and a classical sequential backward feature selection was used by Qu and Zuo [14] to distinguish four different impeller damages and their four different wear degrees, each of which corresponded to a damage mode. Similarly, Qu and Zuo [15] used least square support vector regression for the quantitative evaluation of impeller health condition. Zhao et al. [16] designed a revised neighbourhood rough set model for the selections of useful features and applied them to identify different impeller faults. Then, the joint method of the half, full spectra and principle component analysis was used to find a monotonic performance degradation trend. One of the disadvantages of the above methods used for impeller health condition evaluation is that the data collected from experiments with artificial impeller damages may be less to naturally reflect the true wear evolution of impellers. Recently, the industrial impeller wear data, which was collected by one of the co-authors, was analyzed by Maio et al. [17] using the combination of fuzzy c-means and hierarchical trees. Even though the wear function built was attractive, only some pieces of data were used to validate their method. Through the above analyses, it is concluded that run to alert or failure impeller data

must be investigated to build more natural models for impeller performance degradation assessment. In this paper, we focus on impeller performance degradation assessment using run to alert data. A moving-average mean wear degradation index and a moving-average deviation wear degradation index are proposed to track the health condition of the impellers.

2 Two Health Indicators for Impeller Performance Degradation Assessment

2.1 Proposed Health Indicators for Impeller Performance Degradation Assessment

Vibration analysis is widely used in machine fault diagnosis and prognosis [18]. In order to build health indicators for impeller performance degradation assessment, the vibration components of slurry pumps must be investigated firstly. In this paper, the data was collected from an industrial oil sand pump by one of the co-authors. The oil sand pump was driven by a motor with the rotation frequency equal to f_{m} = 26 Hz. Because the oil sand pump was stepped down through a gearbox, the pump rotation frequency f_{p} was approximately calculated as 6.62 Hz. The vane-passing frequency f_{vpf} was that the number of blades on the impeller (4 blades used in this paper) multiples the pump rotation frequency and it was approximately equal to 26.48 Hz. The gear meshing frequency f_{gmf} was approximately equal to 362 Hz. Because the data was influenced by many unknown factors, the calculated theoretical frequencies may vary over time and not be fixed.

The pump vibration measurements were collected by using the smart asset management system (SAMS) software developed in the Smart Engineering Asset Management Laboratory. The data acquisition equipment consisting of a National Instrument (NI) DAQ 9172 and a DAQ module NI 9234 was used. Four accelerometers were mounted at four different locations of the slurry pump, which were shown in Fig. 1, where the PCB 352A60 accelerometers (S1 and S2) were mounted on 'casing lower' and 'casing discharge' and the PCB 352C18 accelerometers (S3 and S4) were mounted on the 'suction and discharge pipes'. The data was recorded from March to June and the total number of the vibration measurements N was 1,096. In order to identify each measurement, these measurements were numbered by document numbers from 1 to 1,096. The sampling frequency was set to 51200 Hz. For each measurement, the vibration signal with the length L equal to 51200 samples was collected. The data collected from the suction pipe was used for the analyses in this paper.

Assume that N successive slurry pump vibration measurements are denoted as $y_k(t)$, $k = 1, 2, \ldots, N$. Considering the influence caused by unknown outside factors, the vibration measurements are normalized by:

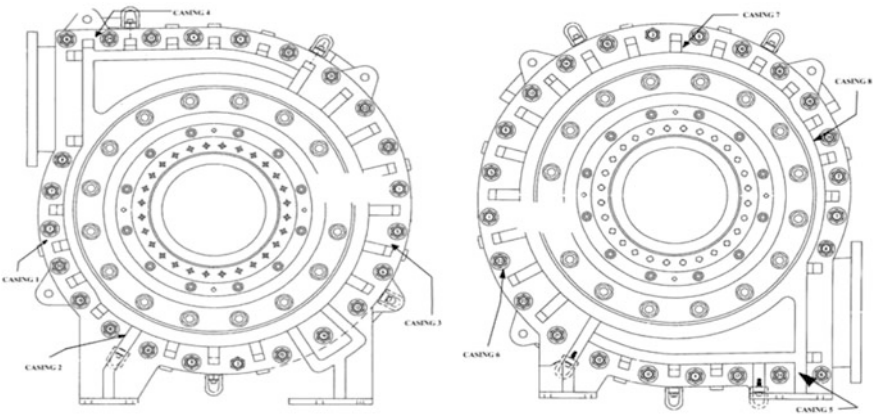


Fig. 1 The locations of the accelerometers used in the oil sand pump

$$y_k(t)=(y_k(t)-\frac{\sum_{t=1}^{L} y_k(t)}{L}) \Bigg/ \sqrt{\frac{\sum_{t=1}^{L} (y_k(t)-\frac{\sum_{t=1}^{L} y_k(t)}{L})^2}{L-1}},\ k=1,2,\ldots,N. \tag{1}$$

The fast Fourier transform of Eq. (1) is given as:

$$y_k(f)=\sum_{t=1}^{L} y_k(t)\mathrm{e}^{-2\pi \mathrm{i}\times(t-1)\times(f-1)/\mathrm{L}},\ k=1,2,\ldots,N. \tag{2}$$

It is known that the unbalance and the misalignment of the motor are related with the motor rotating frequency f_{m} and its harmonics. The sidebands around gear meshing frequency f_{gmf} can be regarded as gear wear. Besides, pumps usually have strong vibration components at vane-passing frequency f_{vpf}. The frequency spectra of the oil sand pump measurements at three different documents 33, 338 and 561 are plotted in Figs. 3a, b and c.

Considering the fact that the vane-passing frequency may vary over time, in this paper, the frequency amplitude summation of the frequency band covering the vane-passing frequency is selected as a fault feature for impeller performance degradation assessment. The rectangles with the dotted lines are the frequency bands covering the vane-passing frequency f_{vpf}, which are highlighted in Fig. 2. The fault feature is called as energy evolution (EE) and its mathematical definition is given as:

$$\mathrm{EE}(k)=\sum_{f=f_1}^{f_2}\sum_{k-K+1}^{k}\frac{y_k(f)}{K},\ k=K,K+1,\ldots,N \tag{3}$$

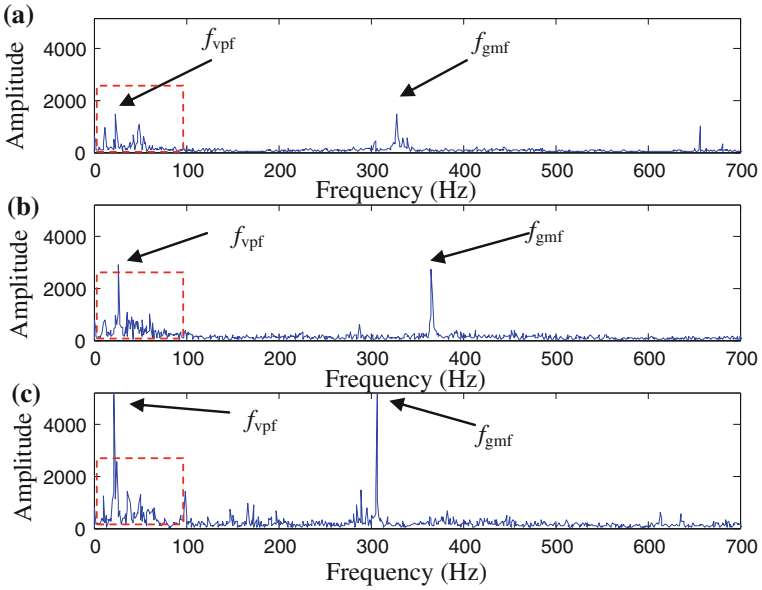


Fig. 2 Frequency spectra of some vibration measurements: **a** at inspection document number 50; **b** at inspection document number 483; **c** at inspection document number 1,050

where f_1 and f_2 are the lower and higher cut-off frequencies. K is the moving-average number. In this paper, f_1 and f_2 are artificially chosen as 20 Hz and 80 Hz. The reasons are given as follows. Firstly, because the frequency resolution is equal to 1 Hz, it is difficult to distinguish the motor rotation frequency with the vane-passing frequency. Secondly, because those frequencies vary over time, it is more reasonable to choose a frequency band rather than some outstanding peaks. The chosen frequency band contains enough fault signatures for impeller performance degradation assessment. In Fig. 3, the evolution of the EE over time is plotted. From the result shown in Fig. 3, it is found that the EE is small at the beginning and

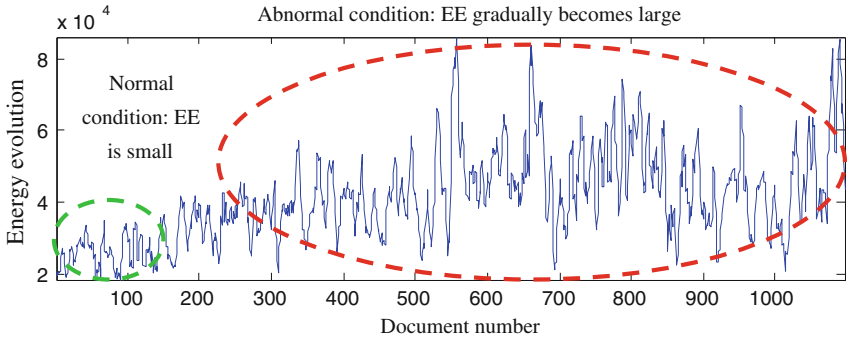


Fig. 3 Impeller health evaluation by using the energy evolution. (The parameter K was set to 5)

then gradually become large and more fluctuated as the document number increases. The parameter K was artificially chosen as 5. The influence of K will be discussed later.

In order to track the underlying trend of the EE, the EE is decomposed into two parts. The first part is the mean of the EE. The second part is the deviation of the EE. Therefore, two health indicators, a moving-average mean wear degradation index (MAMWDI) and a moving-average deviation wear degradation index (MADWDI), are proposed as follows:

$$\mathrm{MAMWDI}(j) = \log\left(\frac{\sum_{k=K}^{j} \mathrm{EE}(k)}{j-K+1}\right) = \log\left(\frac{\sum_{k=K}^{j}\sum_{f=f_1}^{f_2}\sum_{k-K+1}^{k}\frac{y_k(f)}{K}}{j-K+1}\right),\ k = K, K+1, \ldots, N, \tag{4}$$

$$\mathrm{MADWDI}(j) = \log\left(\sqrt{\frac{\sum_{k=K}^{j}\left(\sum_{f=f_1}^{f_2}\sum_{k-K+1}^{k}\frac{y_k(f)}{K} - \frac{\sum_{k=K}^{j}\sum_{f=f_1}^{f_2}\sum_{k-K+1}^{k}\frac{y_k(f)}{K}}{j-K+1}\right)^2}{j-K}}\right), k = K, K+1, \ldots, N \tag{5}$$

The MAMWDI and the MADWDI are plotted in Fig. 4a and b. From the result shown in Fig. 4a, it is seen that Eq. (4) provides the central tendency estimate of the EE. Figure 4b shows the deviation of the central tendency.

It should be noted that the performance degradation assessment only becomes meaningful after the abnormal condition of the impeller is detected. In Fig. 4a, it is found that the MAMWDI only gradually increases after the document number exceeds 100. Therefore, the performance degradation of slurry pump impeller is regarded to start at document number 100.

2.2 *Discussion of Parameter Selection of EE and Two Health Indicators*

According to Eqs. (3–5), it is seen that these equations have the same parameter K. In this section, the parameter K is chosen as different values to show the influence of the different parameter K on the EE and two health indicators. The parameter K was set to 1, 10 and 15, respectively. When the parameter K was set to 1 (it means that there is no moving-average operation), the EE and the two health indicators are plotted in Fig. 5.

When the parameter K was respectively set to 10 and 15, the corresponding results are plotted in Figs. 6 and 7. Compared the result shown in Fig. 5a with the results shown in Figs. 3, 6a and 7a, it is obvious to find that as the parameter K increases, the EE was gradually smoothed. The EE obtained by a larger parameter

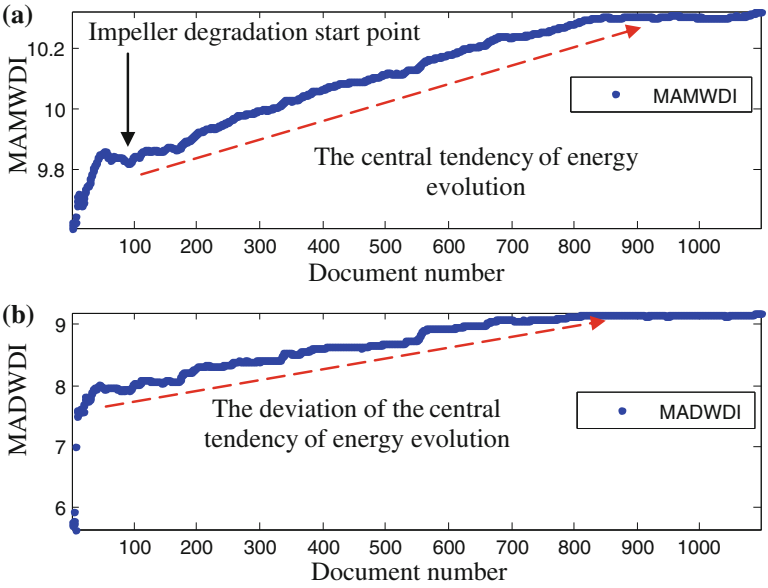


Fig. 4 Impeller performance degradation by using: **a** the moving-average mean wear degradation index; **b** the moving-average deviation wear degradation index. (The parameter K was set to 5)

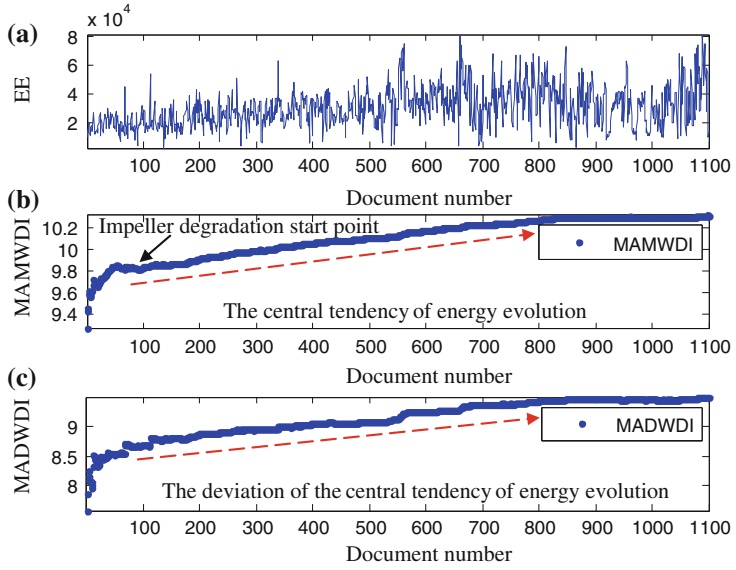


Fig. 5 Impeller performance degradation by using: **a** energy evolution; **b** moving-average mean wear degradation index; **c** moving-average deviation wear degradation index. (The parameter K was set to 1)

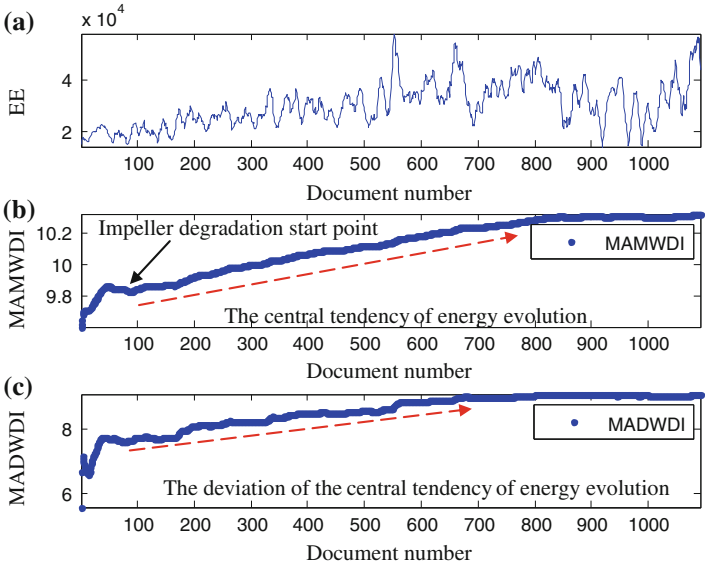


Fig. 6 Impeller performance degradation by using: **a** energy evolution; **b** moving-average mean wear degradation index; **c** moving-average deviation wear degradation index. (The parameter K was set to 10)

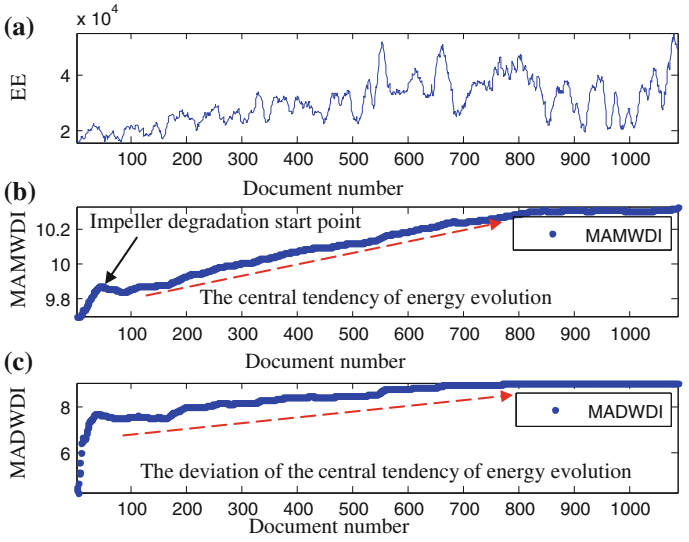


Fig. 7 Impeller performance degradation by using: **a** energy evolution; **b** moving-average mean wear degradation index; **c** moving-average deviation wear degradation index. (The parameter *K* was set to 15)

K have a clearer trend. However, it should be noted that, as the parameter K increases, the computing time of EE also increases.

Therefore, it does not mean that the larger the parameter K is, the better the EE is. The trade-off between the benefit of the smoothness of the EE and the computing cost should be simultaneously considered. Another phenomenon is observed. Even through different parameters were used in the two health indicators, the trends of the two health indicators are not significantly different. Therefore, it is believed that the different parameters have little influence on the two health indicators used for impeller performance degradation assessment.

3 Conclusions

In this paper, two health indictors for slurry pump impeller performance degradation assessment were developed. Firstly, pump vibration components were analyzed. Based on the analyses, low-frequency vibration components were used as a fault feature. Energy evolution was proposed to track the health condition of slurry pump impeller. In order to track the underlying trend of energy evolution, energy evolution was decomposed into two parts: the mean of energy evolution and the deviation of energy evolution. Considering the two parts, a moving-average mean wear degradation index and a moving-average deviation wear degradation index were proposed accordingly. Because energy evolution, the moving-average mean wear degradation index and the moving-average deviation wear degradation index had the same parameter K (the number of moving-average), the influence of the parameter K on the energy evolution, the moving-average mean wear degradation index and the moving-average deviation wear degradation index was investigated. It was concluded that, as the parameter K increased, the energy evolution was smoothed but its corresponding the computing time used for calculating the energy evolution increased. Therefore, the trade-off between the benefit of the smoothness and the computing cost should be considered when the energy evolution was used. Another conclusion is that the different parameters have little influence on the two health indicators used for impeller performance degradation assessment.

Acknowledgment This article was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 122011) and a grant from City University of Hong Kong (Project No. 7008187).

References

1. Qiu H, Lee J, Lin J et al (2003) Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. Adv Eng Inform 17:127–140
2. Wang D, Miao Q, Kang R (2009) Robust health evaluation of gearbox subject to tooth failure with wavelet decomposition. J Sound Vib 324:1141–1157

3. Ocak H, Loparo KA, Discenzo FM (2007) Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: a method for bearing prognostics. J Sound Vib 302:951–961
4. Miao Q, Wang D, Pecht M (2010) A probabilistic description scheme for rotating machinery health evaluation. J Mech Sci Technol 24:2421–2430
5. Hong H, Liang M (2009) Fault severity assessment for rolling element bearings using the Lempel-Ziv complexity and continuous wavelet transform. J Sound Vib 320:452–468
6. Pan Y, Chen J, Guo L (2009) Robust bearing performance degradation assessment method based on improved wavelet packet–support vector data description. Mech Syst Signal Pr 23:669–681
7. Wang D, Tse PW, Guo W et al (2011) Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis. Meas Sci Technol 22:025102
8. Shen Z, He Z, Chen X et al (2012) A monotonic degradation assessment index of rolling bearings using fuzzy support vector data description and running time. Sensors 12:10109–10135
9. Zhu X, Zhang Y, Zhu Y (2013) Bearing performance degradation assessment based on the rough support vector data description. Mech Syst Signal Pr 34:203–217
10. Pan YN, Chen J, Dong GM (2009) A hybrid model for bearing performance degradation assessment based on support vector data description and fuzzy c-means. P I Mech Eng C-J Mec 223:2687–2695
11. Yu J-B (2011) Bearing performance degradation assessment using locality preserving projections. Expert Syst Appl 38:7440–7450
12. Miao Q, Tang C, Liang W et al (2012) Health assessment of cooling fan bearings using wavelet-based filtering. Sensors 13:274–291
13. Wang Y, Zuo MJ, Fan X (2005) Design of an experimental system for wear assessment of slurry pumps. In: Proceedings of the Canadian engineering education association, 2005 Canada. pp 1–8
14. Qu J, Zuo MJ (2010) Support vector machine based data processing algorithm for wear degree classification of slurry pump systems. Measurement 43:781–791
15. Qu J, Zuo MJ (2012) An LSSVR-based algorithm for online system condition prognostics. Expert Syst Appl 39:6089–6102
16. Zhao XM, Hu QH, Lei YG et al (2010) Vibration-based fault diagnosis of slurry pump impellers using neighbourhood rough set models. P I Mech Eng C-J Mec 224:995–1006
17. Maio DF, Hu J, Tse P et al (2012) Ensemble-approaches for clustering health status of oil sand pumps. Expert Syst Appl 39:4847–4859
18. Beebe R (2004) Predictive maintenance of pumps using condition monitoring, Elsevier Science

Application of Maximum Correlated Kurtosis Deconvolution on Rolling Element Bearing Fault Diagnosis

Haitao Zhou, Jin Chen and Guangming Dong

Abstract Maximum correlated kurtosis deconvolution (MCKD) searches for an optimal set of filter coefficients to enhance the periodic impulses by introducing correlation to kurtosis. This method can realize the feature extraction and the diagnosis of rolling element bearing's faults by improving signal to noise ratio (SNR) of signal. In order to obtain a better result, how to select the important parameters of MCKD is discussed in this chapter. After selecting proper parameters, this method is applied to both simulated and experimental data. The result of simulated data shows that this method has potentials in fault diagnosis of rolling element bearing. The experimental data from an accelerated life test of rolling element bearing are used for validation, which shows that this method can successfully detect the incipient fault.

Keywords Correlated kurtosis · Filter · Rolling element bearing · Maximum correlated kurtosis deconvolution

1 Introduction

Rolling element bearings are vital components in many rotating machines. Their failure may lead to great economic loss and threaten people's life. Therefore, detecting and diagnosing bearing faults are very important in machinery condition monitoring and diagnostics [1].

A localized defect will affect vibration signal. In general, when bearing ball strikes the defect, an impulse occurs and the resonance of structure is excited. But if the defect is small, the fault signal is not obvious and submerged in the strong background noise. It is indeed a great challenge to detect the weak fault [2].

H. Zhou (✉) · J. Chen · G. Dong
State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai, China
e-mail: zhouht1988@163.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_16

Different methodologies have been developed to solve this problem and proved to be effective. Spectral kurtosis (SK) [3, 4] is used to design a filter by locating the resonance frequency band, and proved to be very effective in detecting bearing fault. Wigner Ville spectrum [5] and adaptive Wiener filter [6], which are based on cyclostationary properties, are used to realize bearing fault diagnosis by extracting the bearing fault feature. S-SVDR (STMS based SVD method using SVR) [7] has been proved to have a good local identification capability in rolling element bearing fault. Stochastic resonance [8, 9], a nonlinear technique, is also used to enhance the weak periodic impulse.

The minimum entropy deconvolution (MED) technique, proposed by Wiggins [10] in 1978, designs a FIR filter to remove the effect of transmission path by minimizing the entropy of signal. It's effective in seismic survey. Endo and Randall proposed ARMED (AR method followed by MED) [11] to enhance the gear tooth fault signal. Sawalhi et al. [12] demonstrated that ARMED followed by SK is also effective in detecting ball element bearing faults. Based on MED, McDonald proposed MCKD by introducing correlated kurtosis in [13]. Compared with MED, MCKD takes advantage of the periodic nature of the faults as well as the impulse-like vibration behavior associated with most types of faults. It has been demonstrated MCKD performs better than MED in detecting gear chip fault.

This chapter presents the algorithm of MCKD and discusses the effect of main parameters such as order shift and filter length on detecting bearing faults. The signal is filtered by MCKD filter, and then the envelope spectrum based on Hilbert transform is calculated, from which the effectiveness of MCKD technique is demonstrated on detecting bearing faults.

This chapter is organized as follows. The important deconvolution norm CK is presented in Sect. 2. Then the detail of MCKD algorithm is presented in Sect. 3. The effects of order shift and filter length on detecting bearing fault are analyzed in Sect. 4. The simulated deconvolution results of bearing are presented in Sect. 5, which succeed in detecting inner and outer race fault of bearing. An accelerated life test of rolling element bearing is presented in Sect. 6, where the incipient inner race fault is detected and diagnosed successfully. The conclusion is presented in Sect. 7.

2 Correlated Kurtosis

In order to take advantage of periodicity of fault, a new norm is proposed [13] as follows:

$$\begin{aligned}\text{Correlated Kurtosis of first } - - - \text{ shift } &= CK_1(T) = \frac{\sum_{n=1}^{N}\left(y_n y_{n-T}\right)^2}{\left(\sum_{n=1}^{N} y_n^2\right)^2} \\ \text{Correlated Kurtosis of M } - - - \text{ shift } &= CK_M(T) = \frac{\sum_{n=1}^{N}\left(\prod_{m=0}^{M} y_{n-mT}\right)^2}{\left(\sum_{n=1}^{N} y_n^2\right)^{M+1}} \end{aligned} \quad (1)$$

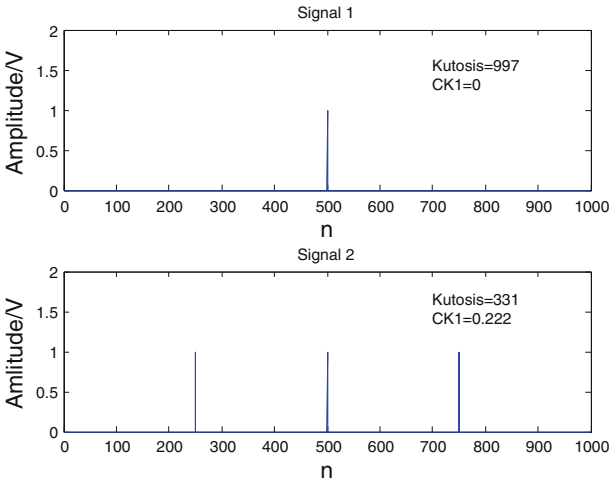


Fig. 1 The kurtosis and correlated kurtosis of two different signals

$$y_n = \sum_{k=1}^{L} f_k x_{n-k+1}, \quad x_n = 0 \text{ and } y_n = 0 \text{ for } n \neq 1, 2, \ldots, N$$

Where N is the number of samples in the output signal $\vec{y}$, L is the length of filter $\vec{f}$, and T is the period of interest. The property of CK differs from kurtosis a lot when the period of interest T is considered.

Figure 1 illustrates the CK_1 versus kurtosis for two simple signals. It shows that the CK_1 approaches a maximum for a periodic impulse about the specified period as opposed to the kurtosis which tends to a maximum with a single impulse.

3 Maximum Correlated Kurtosis Deconvolution

The maximum correlated kurtosis deconvolution technique is a type of system identification method which searches for an optimum set of filter coefficients $\vec{f} = [f_1, f_2, \ldots f_L]$ to recover the desired input signal. The filter design aims at the maximum value of correlated kurtosis in output signal. The process of MCKD is shown as Fig. 2.

The algorithm is researched by McDonald [13]. It's very clear in his paper. The main content is listed as follows:

The objective of the algorithm is to find the filter coefficients which maximize the correlated kurtosis of the output signal y:

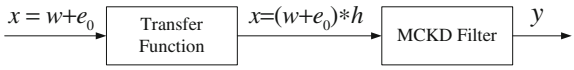


Fig. 2 Inverse filtering (deconvolution) process of MCKD

$$Obj_M(T) = \max_f CK_M(T) = \max_f \frac{\sum_{n=1}^{N} \left(\prod_{m=0}^{M} y_{n-mT}\right)^2}{\left(\sum_{n=1}^{N} y_n^2\right)^{m+1}} \tag{2}$$

The output signal can be described as a convolution as follows:

$$y(n) = x(n) * f_0(n) \tag{3}$$

Optimum filter settings occur when the objective function $Obj_M(T)$ achieves:

$$\frac{dCK_M(T)}{d\vec{f}} = 0 \tag{4}$$

Combining Eqs. (2), (3) and (4), the final equation can be expressed as

$$\vec{f} = \frac{\left\|\vec{y}\right\|^2}{2\left\|\vec{\beta}\right\|} (X_0 X_0^T)^{-1} \sum_{m=0}^{M} X_{mT} \vec{\alpha}_m \tag{5}$$

where $\vec{\alpha}_m = \begin{bmatrix} y_{1-mT}^{-1}(y_1^2 y_{1-T}^2 \cdots y_{1-MT}^2) \\ y_{2-mT}^{-1}(y_2^2 y_{2-T}^2 \cdots y_{2-MT}^2) \\ \vdots \\ y_{N-mT}^{-1}(y_N^2 y_{N-T}^2 \cdots y_{N-MT}^2) \end{bmatrix}$, $\vec{\beta} = \begin{bmatrix} y_1 y_{1-T} \cdots y_{1-MT} \\ y_2 y_{2-T} \cdots y_{2-MT} \\ \vdots \\ y_N y_{N-T} \cdots y_{N-MT} \end{bmatrix}$ and

$$X_r = \begin{bmatrix} x_{1-r} & x_{2-r} & \cdots & x_{N-r} \\ 0 & x_{1-r} & \cdots & x_{N-1-r} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{N-L-r+1} \end{bmatrix}_{L \times N}$$

In order to obtain the filer coefficients, iterative process should be performed as follows:

Step 1: Set the interested period of impulse $T$and the order shift$M$.
Step 2: Assume the initial value of the inverse filter coefficients $f^{(0)}$.
Step 3: Compute the output $y^{(0)}$ using the filter coefficients $f^{(0)}$ and input signal x with Eq. (3).
Step 4: Update the new filter coefficients $f^{(1)}$ with Eq. (5).
Step 5: Compute the error criterion of the algorithm:

$$err = CK_M^{(k)} - CK_M^{(k-1)} \tag{6}$$

If '$err > tolerance$', the iterative process continues and the filter coefficients update by repeating the process from Step 3. If '$err \leq tolerance$', the iteration finishes. Then the filter coefficients are what we expect. The effect of order shift and filter length are discussed in the following section.

4 Parameter Discussion

This section discusses the effects of order shift and filter length on detecting bearing fault through simulations. A famous roller bearing's model based on single point defect, proposed by Mcfadden [14], integrates the effects of rolling bearing geometry, shaft speed, load distribution, transfer function, the decaying exponential, and so on.

$$\begin{cases} x(t) = \sum_{i=1}^{N} A_i \cdot \mathrm{s}(t - iT - \tau_i) + n(t) \\ A_i = A_0 \cos(2\pi Qt + \phi_A) + C_A \\ s(t) = e^{-Bt} \cdot \sin(2\pi f_n t + \phi_w) \end{cases} \tag{7}$$

where A_i is the amplitude modulation with a period of $1/Q$, $s(t)$ is the oscillating impulse with the average inter-arrival time T between two adjacent impacts, τ_i is the tiny fluctuation around T, $n(t)$ is a white stationary noise, C_A is an arbitrary constant, B is the damping coefficient depending on the system, and f_n is the natural frequency of the system.

4.1 Effect of Order Shift M

Since the objective function $CK_M(T)$ relies on the order shift, it's necessary to analyse the order shift's influence. Higher order shift requires better estimates of the fault period T. However, due to the existence of random slip of rolling elements, the impact periodicity is not exact. From the above, M cannot be too high in detecting the bearing fault because of the fluctuation τ_i. The parameters of simulated bearing outer race fault signal are as follows: Sampling rate 25.6 kHz, fault frequency 60Hz. The relation between $CK_M(T)$ and $Var(\tau)$ (the variance of τ_i) is shown in Fig. 3. The value of $CK_M(T)$ is sensitive to the variance of τ_i. It decreases sharply as

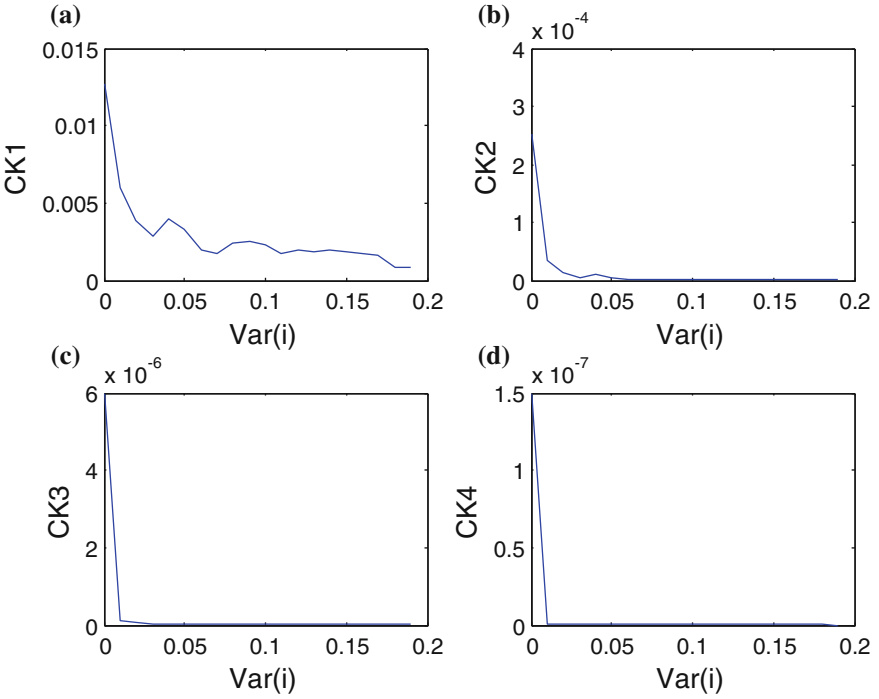
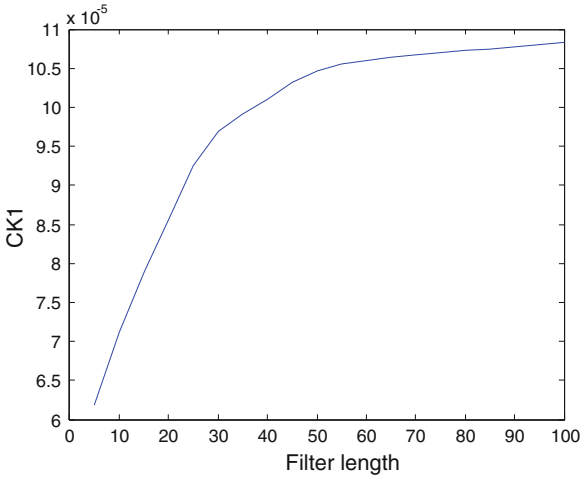


Fig. 3 The relation between correlated kurtosis of M shift and variance of τ_i: **a** M = 1. **b** M = 2. **c** M = 3. **d** M = 4

Fig. 4 Filter length's effect on correlated kurtosis of the filtered signal with MCKD



$Var(\tau)$ increases. And for higher order shift, it decreases more quickly than for the lower one. Therefore, higher order shift doesn't benefit the optimization of $CK_M(T)$ in bearing signal. Therefore, the order shift $M = 1$ is suggested.

4.2 *Effect of Filter Length* L

The filter length is also an important parameter in filter design. The correlated kurtosis of first-order $CK_1(T)$ in the output signal can be used to select proper filter length. Here the same simulated signal as in Sect. 4.1 is presented here to analyze the relation between $CK_1(T)$ and filter length$L$.

Figure 4 shows that $CK_1(T)$ increases as the filter length L does. After the filter length reaches 60, its influence on output signal seems not so obvious. Therefore, the filter length $L = 100$ is suggested.

5 Simulation Analysis

The bearing signals with inner and outer race defect are simulated by Eq. (7). The sampling frequency f_s is 25.6 kHz, the natural frequency f_n is 4 kHz, the shaft rotational frequency f_r is 10 Hz, the inner race fault frequency f is 65 Hz, the outer race fault frequency f_o is 47 Hz, the data length N is 25,600, the random slip τ_i is 0.01, and the SNR = −12 dB. The order shift M is 1 and the filter length L is 100 as suggested in Sect. 4.

The simulated bearing signal with inner race fault is given in Fig. 5. It shows the effectiveness of MCKD technique in bearing fault diagnosis. The signal in the top row (a) represents the raw signal without any processing. It is seen that the background noise is so strong that the fault cannot be detected. In the second row (b), the time waveform of filtered signal is presented. After the enhancement of MCKD filter, the periodic fault feature is obvious in the time waveform. The third row (c) represents the envelop spectrum of the filtered signal. It clearly shows that the characteristic frequencies are f_i and its harmonics modulated by f_r, which implies the occurrence of inner race fault. Figure 6 shows that this technique is also effective in detecting outer race fault.

6 Experiment Validation

Electrical discharge machining is used to obtain artificial defects. But it cannot reflect practical machine operation. For validation, vibration signal is collected in a rolling element bearing accelerated life test. Rolling element bearing accelerated life test is performed to collect vibration data over whole life time in Hangzhou Bearing

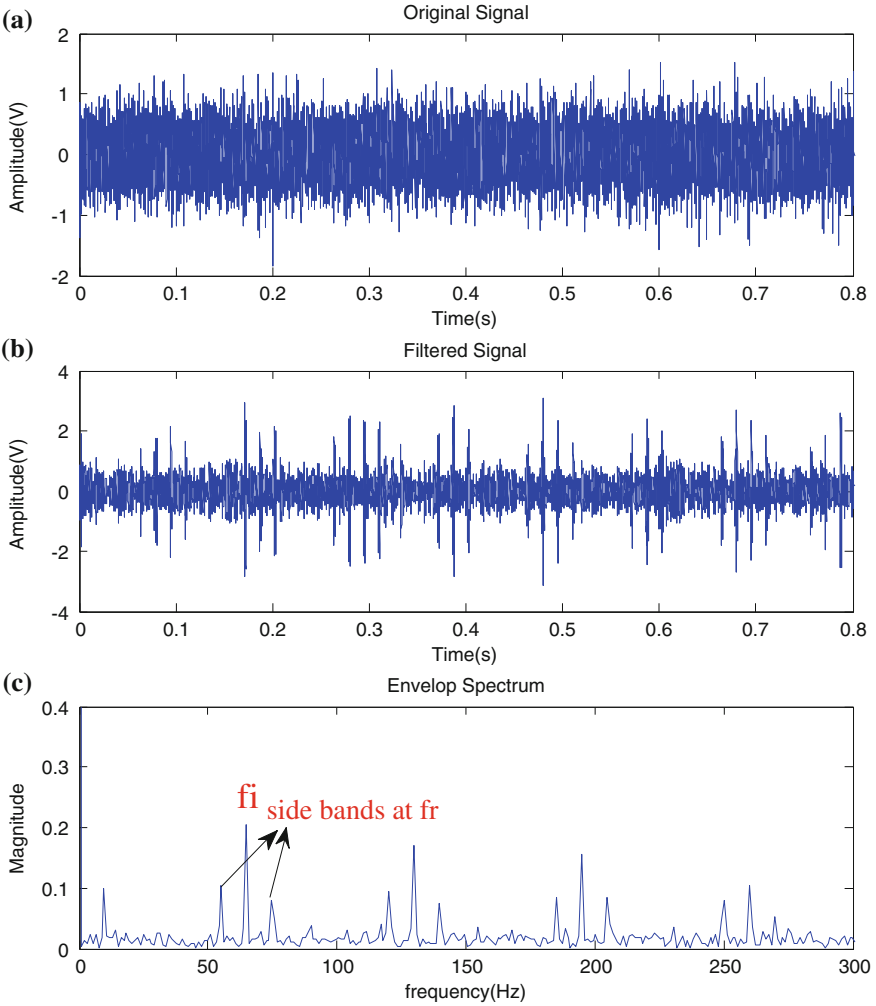


Fig. 5 Simulated signal with inner race fault: **a** original signal. **b** filtered signal. **c** envelop spectrum

Test and Research Centre (HBRC). It simultaneously hosts four rolling element bearings on one shaft driven by an AC motor and coupled by rubber belts. If any bearing is failed, a new one will replace it. The experiment rig is shown in Fig. 7. Four same rolling element bearings are tested and their type is 6307 with its corresponding parameters and operating condition shown in Table 1. Five characteristic frequencies are presented in Table 2. The data acquisition system includes three acceleration transducers and DAQCard-6023E. One group of data was

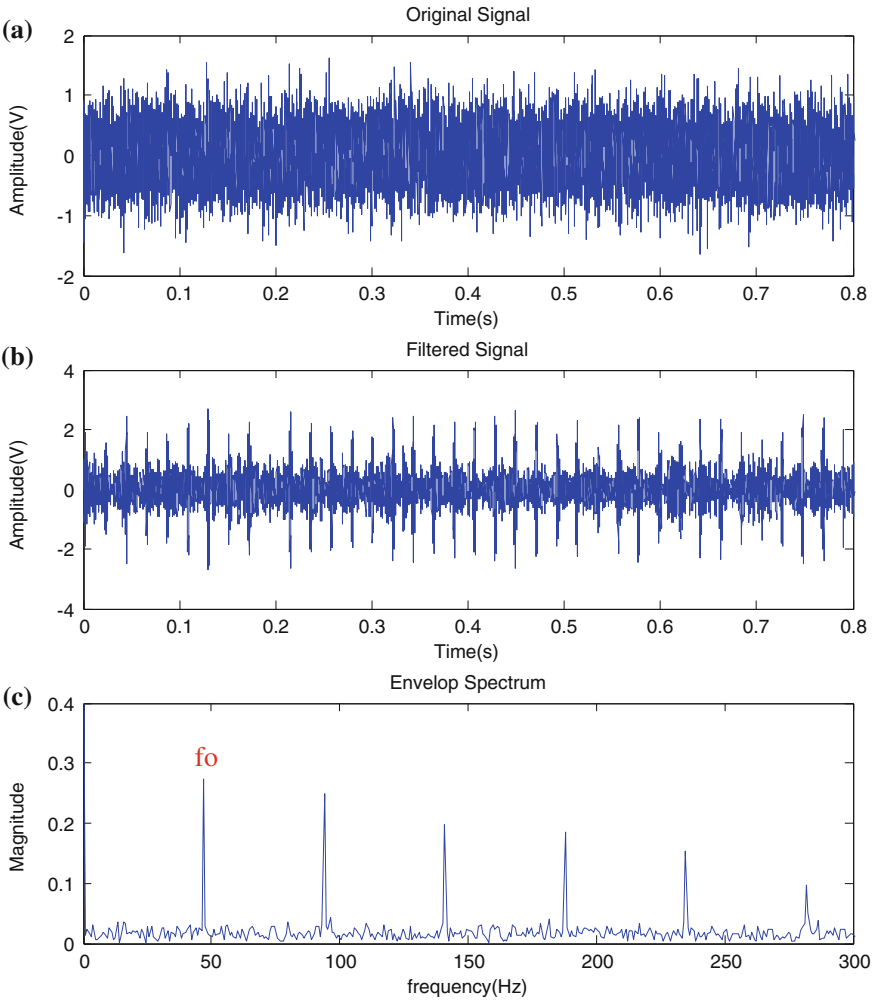


Fig. 6 Simulated signal with outer race fault: **a** original signal. **b** filtered signal. **c** envelop spectrum

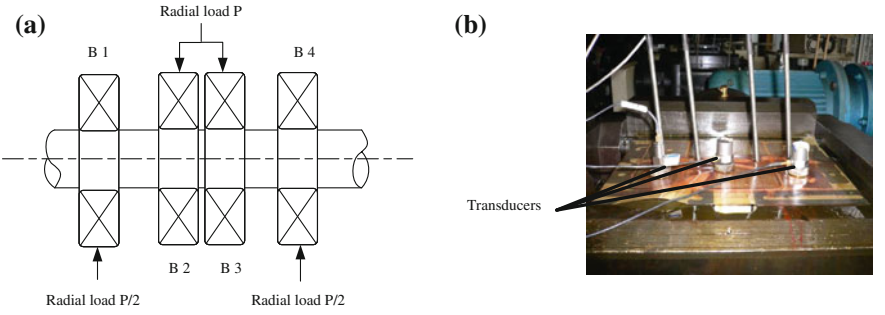


Fig. 7 Accelerated rolling element bearing life test rig, **a** Location of acceleration transducers **b** Experimental test rig

Table 1 Rolling element bearing parameters and operation conditions

Type	Ball number	Ball diameter (mm)	Pitch diameter(mm)	Contact angle	Motor speed (rmp)	Load (KN)
6307	8	13.494	58.5	0	3000	12.744

Table 2 Main feature frequency

f_r	f_c	f_b	f_i	f_o
50	19	102	246	153

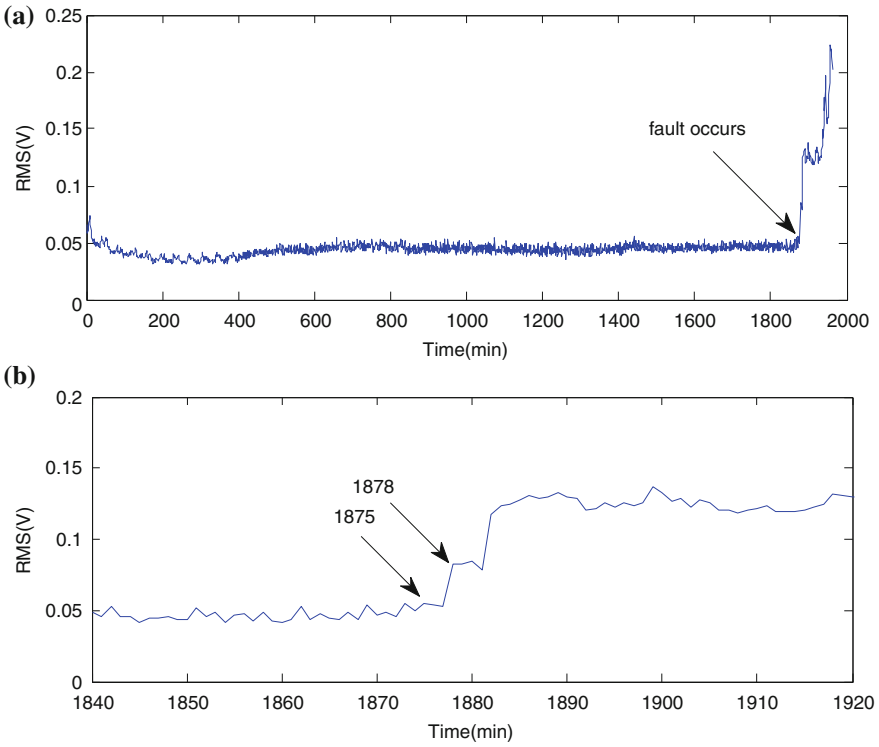


Fig. 8 The RMS during the bearing's whole lifetime, **a** RMS of the entire life, **b** RMS localization around 1878th min

collected every minute with the data sampling rate of 25.6 kHz and the data length is 20,480 points.

The RMS is one of the most widely used indices to monitor the bearing condition. Figure 8 represents the RMS over the entire life time of B3. There are 1962 groups of data. It is seen that the RMS starts to increase at the group 1878. When

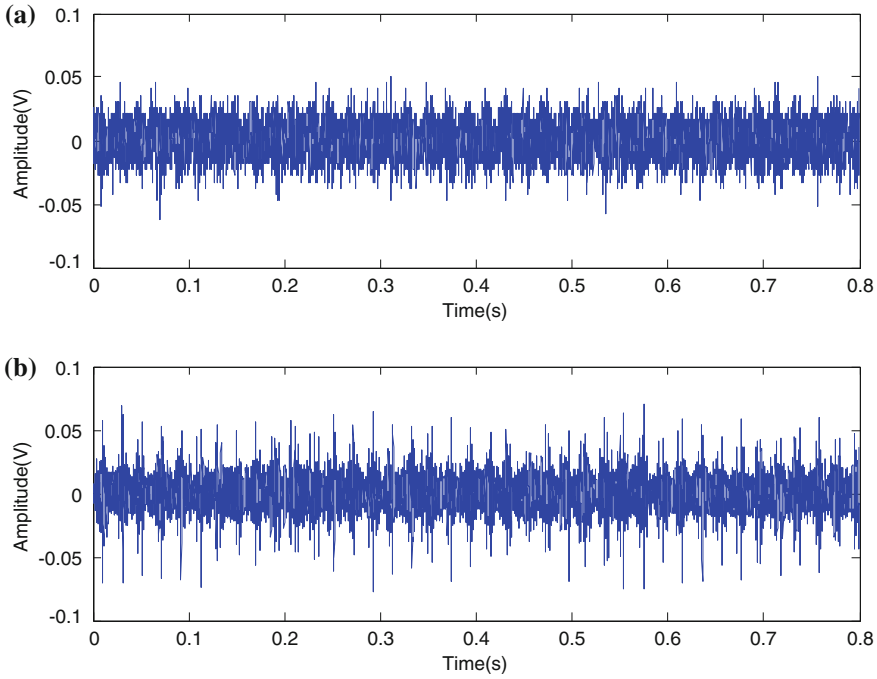


Fig. 9 Time waveform of the signal: **a** original signal. **b** filtered signal with MCKD

the experiment rig goes on running, the RMS continues to increase sharply. That means there must be certain fault in the B3. The test rig was stopped to avoid the machinery damage. So it's very significant to detect the incipient defect in bearing condition monitoring. The proposed data group 1875, earlier than 1878, is used to extract fault feature by MCKD.

Figure 9 represents the time waveform of original and filtered signal in group 1875. In the time waveform of original signal, there is no evident fault feature. But abundant periodic impulses are obvious in the filtered signal. It can be inferred that B3 suffers from certain fault. It's necessary to give further analysis to confirm the fault type. Figure 10 represents the PSD estimation of the original and filtered signal. The envelop spectrum of both original and filtered signal is shown in Fig. 11. The inner fault frequency f_i and its sideband f_r are much more obvious in filtered signal than the original one.

After performing the MCKD technique, the bearing is diagnosed as inner race fault. For validation of the analysis, the fault bearing is examined carefully. Figure 12 shows the failed bearing. In conclusion, the MCKD technique is effective in detecting the incipient fault under strong background noise.

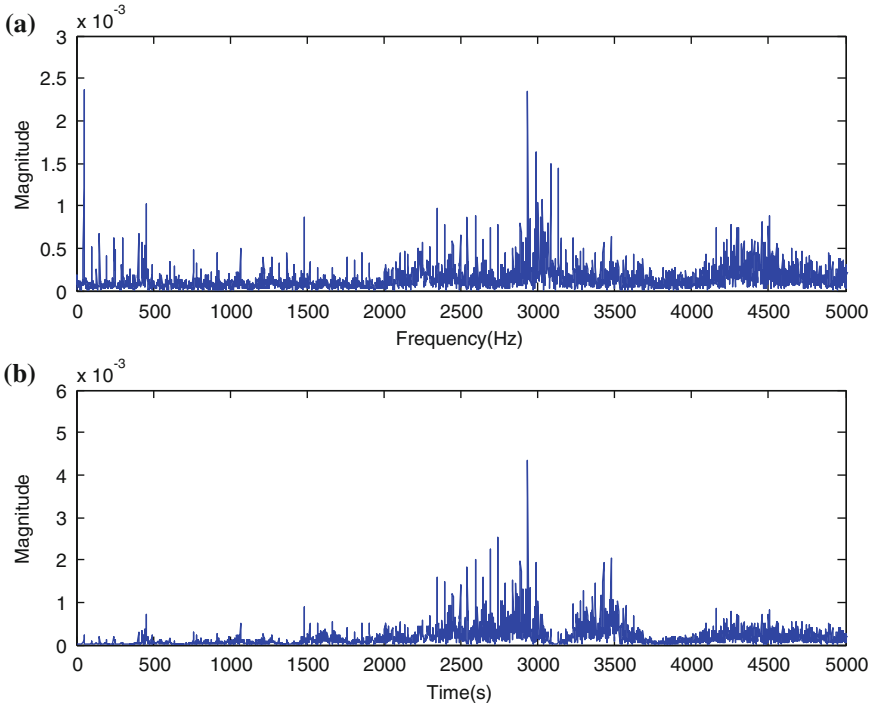


Fig. 10 PSD estimation of the **a** original signal, and **b** filtered signal with MCKD

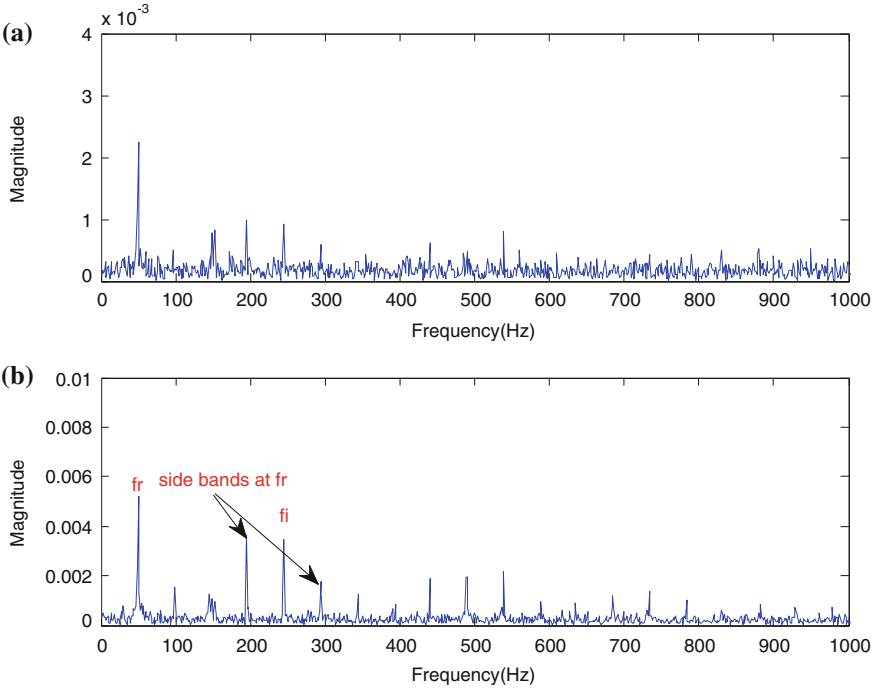


Fig. 11 Envelop spectrum analysis of the **a** original signal, and **b** filtered signal with MCKD

Fig. 12 The photos of failure bearing (inner race fault)

7 Conclusion

This chapter presents the new deconvolution process, MCKD, which searches for an optimal set of filter coefficients to enhance the periodic impulses by introducing correlation to kurtosis. The vital parameters, order shift and filter length, are discussed and chosen for detecting the bearing fault. The order shift cannot be too large for its sensitivity to the fluctuation in bearing signal. The simulation data and an accelerated life test prove its validity in rolling element bearing fault diagnosis. The accelerated life test collected the bearing vibration data over the entire lifetime (normal-fault-failure). The RMS shows the bearing fault occurred after it ran for 1878 min. The combination of MCKD and envelop spectrum can detect the incipient fault 3 min earlier than the RMS can. The fact proves the effectiveness of the MCKD technique in rolling element bearing fault diagnosis.

Acknowledgments Support for this work from Natural Science Foundation of China (Approved Grant: 51035007 and 51105243) is gratefully acknowledged. The authors would also like to appreciate the support of Hangzhou Bearing Test and Research Center (HBRC) on the experiment.

References

1. Randall RB, Antoni J (2011) Rolling element bearing diagnostics—a tutorial. Mech Syst Signal Process 25(2):485–520
2. Park C-S, Choi Y-C, Kim Y-H (2013) Early fault detection in automotive ball bearings using the minimum variance cepstrum. Mech Syst Signal Process 38(2):534–548
3. Antoni J (2006) The spectral kurtosis: a useful tool for characterising non-stationary signals. Mech Syst Signal Process 20(2):282–307
4. Antoni J, Randall RB (2006) The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. Mech Syst Signal Process 20(2):308–331
5. Dong G, Chen J (2012) Noise resistant time frequency analysis and application in fault diagnosis of rolling element bearings. Mech Syst Signal Process 33:212–236
6. Ming Y, Chen J, Dong G (2011) Weak fault feature extraction of rolling bearing based on cyclic Wiener filter and envelope spectrum. Mech Syst Signal Process 25(5):1773–1785
7. Cong F et al (2013) Short-time matrix series based singular value decomposition for rolling bearing fault diagnosis. Mech Syst Signal Process 34(1–2):218–230

8. Li J, Chen X, He Z (2013) Adaptive stochastic resonance method for impact signal detection based on sliding window. Mech Syst Signal Process 36(2):240–255
9. Qiang L et al (2007) Engineering signal processing based on adaptive step-changed stochastic resonance. Mech Syst Signal Process 21(5):2267–2279
10. Wiggins RA (1978) Minimum entropy deconvolution. Geoexploration 16(1–2):21–35
11. Endo H, Randall RB (2007) Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter. Mech Syst Signal Process 21(2):906–919
12. Sawalhi N, Randall RB, Endo H (2007) The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. Mech Syst Signal Process 21(6):2616–2633
13. McDonald GL, Zhao Q, Zuo MJ (2012) Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection. Mech Syst Signal Process 33:237–255
14. McFadden PD, Smith JD (1984) Model for the vibration produced by a single point defect in a rolling element bearing. J Sound Vib 96(1):69–82

The Role of Life Cycle Cost in Engineering Asset Management

Khaled El-Akruti, Richard Dwight, Tieling Zhang and Mujbil Al-Marsumi

Abstract This paper presents a case study demonstrating life cycle cost (LCC) analysis as a major and critical activity of engineering asset management decisions and control. The objective is to develop a maintenance policy to control the economics of replacement and repair practice of refractory lining of an electric arc furnace (EAF). The replacement and repair policies involve the optimum life policy, the repair versus replacement policies, the repair limit method and the comparison of lining material types from different suppliers. The developed models provide a method for defining the most important factors involved in decision making with respect to operational and managerial levels. The approach also involves deciding the remaining age value as the repair limit criteria while avoid lining failure due to unavoidable deterioration caused by variation in operation conditions. The decision criteria are established as: (a) what type of lining material is better to use? (b) When to replace lining in a cyclic manner? (c) At what sequence is hot repair required and (d) whether to replace or use cold repair between replacements. Finally, the model output values for the decision criteria are presented in tables and graphs to guide decision making in operation and maintenance.

Keywords Life cycle cost · Engineering asset management · Decision-making criteria · Optimal maintenance · Decision-making model

K. El-Akruti (✉) · R. Dwight · T. Zhang
Faculty of Engineering and Information Science, University of Wollongong, Wollongong, Australia
e-mail: khaled@uow.edu.au

R. Dwight
e-mail: radwight@uow.edu.au

T. Zhang
e-mail: tieling@uow.edu.au

M. Al-Marsumi
Industrial Engineering, Higher Institute of Industry, Misurata, Libya
e-mail: mujbila@hotmail.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_17

1 Introduction

The concept of asset management utilizes the LCC analysis in managing the life of assets. There is a common understanding that Engineering Asset Management; AM involves life cycle management which is based on LCC as a dominant criteria for decision making within the AM system [2]. The AM system is defined as: "The system that plans and controls the asset-related activities and their relationships to ensure the asset performance that meets the intended competitive strategy of the organization" [11]. This definition provides an integrated view of the AM system within the whole organization's management system. As a control system AM involves a set of planning and control activities at different organizational levels.

It is proposed that the role of LCC analysis in AM system is focused on defining decision criteria for the lifecycle management of physical assets as a holistic approach to control the life cycle activities of assets in order to achieve the organization's objectives.

The activities of concern in asset management in relation to LCC analysis during each stage of the asset life are shown in Fig. 1. For example, at the preliminary system design stage, the AM system activities that require LCC analysis may include system definition, system analysis, and evaluation of alternatives or trade-offs. The challenge in managing the entire asset life effectively lies in integrating the fragmented activities through the various stages [7]. This leads to integrating the need-identification, alternative analysis, and project selection to the business management focus (ISO/IEC 15288 [21].

Value creation is a concept that is related to organizational activities and LCC [37]. Relationship between value activities and AM system is not clear but

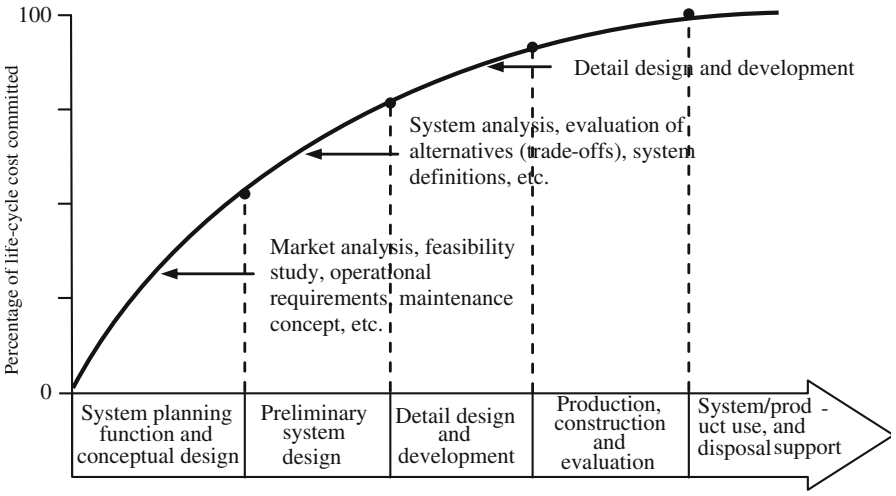


Fig. 1 Life cycle processes and cost committed [4]

performance attributes reflect the LCC-value relationships between activities. El-Akruti [11] argues that the AM system controls LCC of assets. He states, that it also incorporates coordination activities to maintain relationships between these life cycle activities such as those related to:

a. Procurement, finance and accounting which are important for establishing the LCC requirements to enable investment, funding and budgeting, cost analysis and decision making.
b. The information technology to establish the required information flow to facilitate a data base for LCC. For example, a published case study [20] reports that "BP connects its business processes with over 1,500 suppliers to co-ordinate the maintenance, operation and repair of specialized exploration and production equipment".
c. External suppliers to establish their impact on asset's LCC and maintain value added relationships with suppliers and to make outsource verses in-house decision and maintain both-side-benefit relationship.
d. Technical support and development to establish the required development in assets or asset-related processes and the suitable technology or any new developments in technology for use in enhancing performance.
e. Human resources, inventory, quality and safety systems for better performance, less risks and safe environment.

The AM activities involve the use of LCC analysis in lifecycle management of assets in existing organization [1, 7]. Any decision concerning the portfolio of assets is built on the accumulated information of managing the utilisation stage. A major portion of such information would be related to cost of activities at different stages of asset life [4]. At any time it may be determined for example that the current design of one or more assets is not capable of achieving the required performance given the current or projected future environment [9]. Concurrently, organizations must identify the business needs, and make decisions to launch any change or project to enhance assets, their design, operation, maintenance or logistic support [7, 8, 31]. AM projects may involve decision regarding, upgrading, expansion, support system, redesign, replacement or retirements of assets. These decisions require the LCC analysis to choose the right assets, use or maintain them appropriately, and balancing short-term performance against long-term sustainability.

The literature on this topic is extensive. Examples on the use of LCC in asset management decisions are numerous but each example may be unique due to the nature of different assets in different industries. Table 1 is a summary of some of the reviewed literature on the use of LCC in engineering asset management decisions. These decisions highlight the need for LCC analysis as a holistic approach to AM system activities in relation to achieving an organization's objectives. For example, Pinjala et al. [36] discuss relationship between business and some of the asset-related activities such as maintenance.

A strategic approach to maintenance as an asset-related activity has been recognized especially in capital-intensive industries [30, 36, 45].

Table 1 Use of LCC in engineering asset management decisions

Use of LCC in AM decisions	Sample references
To develop value-orientated decision support systems for maintenance and replacement or rehabilitation: this may include optimization or performance improvement, setting policies or developing strategies	Scarf et al. [38], Taylor [43], Tähkämö, et al. [42], Shahata and Zayed [40], Schuman and Brent [39], Khan [26], Mahapatra. [29], Garcia et al. [14], Hartman [18], Eginhard [10], Jardaine [23], White [48] and Herbert and Gordon [19], Wijnia et al. [49]
To evaluate feasibility and/or requirements of existing and new asset or asset-related projects	Goralczyk and Kulczycka [17], Wubbenhorst [51], Buys et al. [5], Vorarat and Al-Hajj [47]
To develop value-orientated decision support systems to improve asset design/selection, installation, use & maintenance and retirement or trade-off between alternatives to an asset	Girsch et al. [16], Barringer [3], Janz and Westkamper [22], Liu [28]; Farran and Zayed [13], Jun and Kim [25]
To assess or assert trade-offs for environmental impact and sustainability	Nyuk et al. [33], Sullivan and Young [41], Norris [32], Mahapatra [29], Castella et al. [6]
To provide a decision tool (e.g. cost-benefit analysis) for estimating project requirements or investment, identify cost drivers and highlight need for change or for selecting assets	Kim et al. [27], Patra and Kumar [35], Ge and Wei [15], Yu-Rong et al. [52], Woodward [50], Jeromin et al. [24], Esveld [12], Thoft-Christensen [44], Uppal [46]

As presented by Blanchard's model [4], integration of LCC models into asset management decision process involves 12 steps:

1. Define system requirements and TPMs.
2. Specify the system life-cycle and identify activities by phase.
3. Develop a cost breakdown structure.
4. Identify input data requirements.
5. Establish costs for each category in the CBS.
6. Select a cost model for analysis and evaluation.
7. Develop a cost profile and summary.
8. Identify high-cost contributors and cause-effect relationships.
9. Conduct a sensitivity analysis.
10. Identify priorities for problem resolution.
11. Identify additional alternatives.
12. Evaluate feasible alternatives and select a preferred approach.

Decisions may involve issues such as:

- Establishing the remaining costs (given you are in the use phase), which raises the issue about replacement cost as a function of behaviour of the current system.
- Repair/replace decision logic which may give rise to economic or optimum repair frequency and replacement period.

- Prediction and estimation decisions which require CBS breadth and depth for visibility.
- Projection decisions which involve investment, system operation and support costs. These are based on the projected activities throughout the operational use and support phase and are usually the most difficult to estimate.
- Trade-off decisions which may involve capital vs. running costs, labour and materials versus reduced services and reduced safety.
- Alternative options decisions which comparing LCC of alternative asset e.g. pieces of equipment or maintenance strategies or methods and balancing the cost of a new item against the cost of maintaining efficiency on the old one and/or that due to the loss of efficiency.

2 LCC Criteria for AM Control Models and Performance

It is necessary at this stage to stress that the purpose of a LCC model is so that it can formally be manipulated to determine relationships between AM control decisions and levels of performance. LCC models are essential for AM control to aim at improving performance, either in terms of improved benefits for the same cost, or reduce cost for the same benefit, or in terms of cost/benefit mixtures. For example for obtaining optimum LCC, control of replacement and/or repair frequencies is needed. AM control is meaningless unless there is a criterion to tell when control is good or bad. Such criteria do not exist, and involve a search for it in the context of the demand for the asset and in the effectiveness of the functions in meeting this demand. It is therefore a two way interaction e.g. replacement and/or repair functions as shown in Fig. 2, and the demand pattern has to be decided in the light of economic or optimum LCC.

As shown in Fig. 3, there are two aspects to this; firstly determination of how to measure performance and secondly identification of the decision criteria (control variables) to be manipulated by the replacement function in order to set a decision or policy. Measures of performance depend on the availability and performance of the asset required by production over some time span. If they are laid down clearly

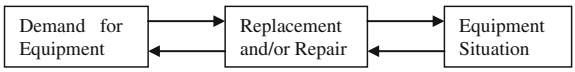


Fig. 2 Demand versus replacement and repair

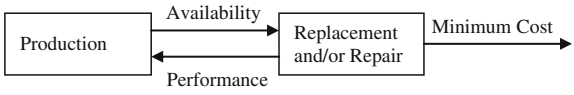


Fig. 3 Production performance versus replacement and repair cost

and can be met, then performance can be judged in terms of cost of providing this service.

Even if availability and performance are specified, one can question the true financial production effects of each service level. Thus what are the financial effects of lost production time and of poor quality? What are the financial effects of delaying production, resulting from not having equipment available? This gives rise to the need of combining both functions; minimizing LCC to achieve various service levels and the economic benefits of various service levels. Finally it needs to be remembered that the way in which an asset is used in production programs, will inevitably influence the condition of the equipment. Meeting demand is usually a priority but some time the choice between delaying production and using asset with risky condition has to be made depending on deterioration rate which may be rapidly accelerated and lead to catastrophe.

Regarding LCC decision criteria for AM policies, the decisions which can be taken are numerous, and they occupy different hierarchical levels, e.g. replacement of equipment, repair or overhaul of equipment, replacement of components or inspection of component. The effect of such decisions cannot be separated. Thus it is the combined effect of operation and maintenance that has to be assessed when dealing with AM policies in utilization stage. For example an optimum overhaul frequency of replacement of some components depends on the frequency of other associated components replacement (component may or may not be replaced at times of major overhauls and at time of breakdown). LCC is needed for replacement decision which does not have to be in the strict sense of the word, but perhaps maintenance decisions such as repair and overhaul may be taken synonymous with replacement provided that it is reasonable to assume that maintenance actions return equipment to as new condition. Therefore it is not a 'one-off' decision, but a serious of decisions. Thus there is a complexity of decision structure that involves LCC.

3 LCC Criteria for AM Control Models and Performance

3.1 Case Study Definition

This case study examines furnaces as the significant assets in steel making industry. In this steel making case study organization, EAFs are the most critical assets for the process availability. The main activities of the EAFs that impact process availability are lining replacement and repair. Therefore, the development of the economical criteria for the control of lining replacement and repair of EAFs provides a good opportunity to increase product unit profit (profit margin).

The policy of lining refractory replacement and/or repair varies greatly from one steel plant to another depending on differences in environmental and operational condition. Hence every plant has to develop its own convenient replacement and/or repair policy.

Furthermore the maintenance strategy of working lining replacement and/or repair controls to a large extent the availability and productivity of the process, resulting in a great effect on the unit cost of liquid steel produced.

The life of the working lining is dependent on the repair practice. Lining repair involves two types; hot repair and cold repair. Hot repair is done by fettling and gunning while EAF is hot. The amount of material used and the time required to do hot repair are the main factors in considering hot repair cost. Cold repair is done by cooling EAF down to fix the damaged spots. It is usually resorted to when deterioration or damage cannot be handled by hot repair. It is time consuming because it requires cooling the EAF down resulting in a great loss of EAF's availability. The time required for cooling EAF and the time required for repairing are the main factors in considering cold repair cost. Working lining replacement is done periodically as required and has direct effect on EAF availability.

Hence it can be concluded that, working lining replacement is directly related to the overall cost of replacement and repair, and lost production cost due to unavailability. Consequently an assessment of working lining replacement costs in terms of lining material and replacement stoppage is required to base decision on the LCC.

Hot repair is done as required and involves consumption of material and relatively short time. Cold repair is done when hot repair is not sufficient and the time is too early for replacement, or most of the lining is still in a good condition. Cold repair is considered as a partial replacement where the EAF has to be cooled down. It involves the use of used bricks and its duration may be as long as that of the replacement or even more.

Hence it can be concluded that, hot and cold repair are as relevant aspects as working lining replacement because they are directly related to LCC involving replacement and repair, and lost production cost due to unavailability. Consequently an assessment of repair costs in terms of hot and cold repair material and hot and cold repair stoppages is required to base decision on the LCC.

3.2 AM Decision Models and LCC Criteria of the Case Study

In this case study, the AM control focuses on providing a maintenance policy that guaranties and outlines the economical decision bases for replacement, repair practice and refractory procurement policy.

For AM control, the main concerns are the working lining life, the required repair amount, frequency and the required repair type. Since the consumption rate of hot repair material and time making hot repair increases with longer life of EAF's lining, there must be a point where it is not worthwhile to use hot repair and replacement is more economical. Furthermore, is it worthwhile to use cold repair at all? Also, what lining supplier's set of refractory materials is more economical?

It is the aim of this paper to demonstrate and establish the modelling procedure using the LCC to set an AM policy that economically control the replacement and

repair practice of refractory working lining in EAFs. The policy of determining the maintenance economical decision bases, involves modelling for the replacement and the repair criteria of lining such that the LCC of lining is reduced for each unit produced. The question then becomes how to optimize LCC while maintaining the same high level of availability, quality and productivity.

As a result, this case study involves answering these questions:

1. What lining type (supplier) is better?
2. When to replace lining in a cyclic manner?
3. At what sequence is hot repair required relative to lining life?
4. Whether to use cold repair in between replacements and determine the repair limit for use?

3.3 Cost Structure Breakdown and Evaluation for Modelling

As the decision criteria are based on cost data, the various types of costs that might be involved in lining repair and replacement must be outlined. These costs are structurally related as shown in Fig. 4. As will be noticed only some costs will be used for the development of the model, where others would not be used either because they have no effect on the replacement and repair practice or they do not change with respect to time or replacement and repair events.

The main cost variables evaluation includes:

1. Working lining cost: which is composed of material cost and stoppage loss and it may be defined as; (Cw = Cwm + Cws). Where Cwm represents the material cost and Cws represent the stoppage loss.
2. Working Lining Replacement Stoppage Cost (Cws): which is a loss of time that could have been utilized for production resulting in two effects one is the loss of operation while incurring ongoing payment of fixed cost. The second is the loss

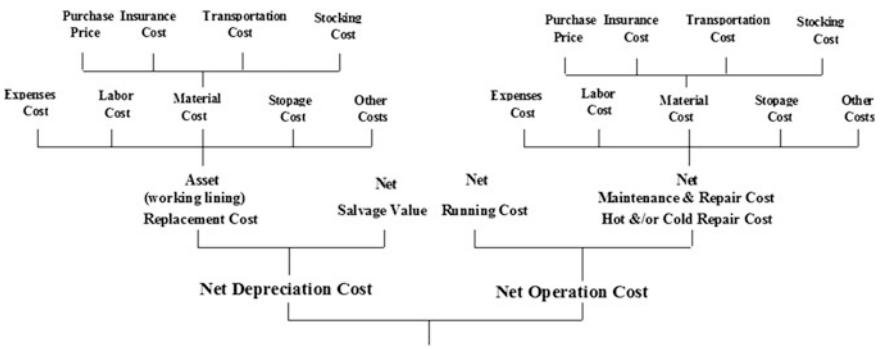


Fig. 4 Model's cost structure breakdown

of products that could have been sold and therefore a possible result of lost benefit. These effects are dependent on many economical policies imposed by strategies of the overall economy. For instance at one extreme, the policy may be purely economical benefit where product value is equal to sales price. On the other extreme the policy may be anything but not economical benefit where product value is equal to product unit cost. For nature of demand in markets associated with this case study it is assumed that stoppage cost is only due to the contribution of "no operation cost" which is incurred ongoing payment of fixed cost.

3. Cold repair cost which is composed of material cost and stoppage loss and is defined as; Cc = Ccm + Ccs. Where Ccm represent the material cost and Ccs represent the stoppage losses. Ccm is the cost made up of the various material used which may be brick or mixes or binding material. Cold repair stoppage cost may be define as:
   Stoppage cost = stoppage duration * productivity * availability * fixed unit cost (Cws = Dcs * Pr * A * NOCu)
   where, Dcs, Pr and A values are to be evaluated statistically from operation records as an input to the model and the NOCu is based on accounting records [34].
4. Hot repair cost which is composed of material and stoppage cost (Ch = Chm + Chs). Where hot repair is defined in terms of gunning and fettling (Chm = Cf + Cg) and the hot repair stoppage cost is define similar to cold repair stoppage cost as: Stoppage cost = stoppage duration * productivity * fixed unit cost (Chs = S * Pr * NOCu).

3.4 Model Development, Result and Application

3.4.1 Model Development

The model is developed based on LCC structure to achieve the research objectives. It is intended to contribute a significant saving in refractory consumption by proper application of the developed model to set the policy for the economical criteria. It is also intended that such application of the model is made such that it does not impose any changes in the actual practice and provides the required information for decision makers to base their decisions on the economical aspects of LCC optimization while increasing productivity and maintaining the required quality. As such those objectives shall be achieved through obtaining an optimum replacement model and a repair limit model.

These models are developed as a decision support system to enhance optimum life decision, suppliers' refractory lining selection decision based on the most economical supplier's lining (refractory set), decision regarding amount and sequence of using hot repair and cold repair limit. The models are developed to define the

appropriate action in terms of hot repair, cold repair or replacement. The two types of repair stand as preventive means to failure and as a substitute for replacement until they become uneconomical. Hot repair is to be carried out as a preventive treatment against failure until optimum life is reached based on minimum LCC per heat. If hot repair fails as a preventive treatment, then cold repair is applied only if it is more economical than replacement. This comparison is done on the basis of the remaining age value of the lining as a repair limit. Cold repair is carried out as long as its estimate at any specified age does not exceed the repair limit.

3.5 Model Results

A graphical presentation of the results for one supplier's material type is presented in Figs. 5 and 6. They show the analysis that determined the criteria for optimising replacement and frequency of hot repair for one lining supplier. These solutions represent the evaluation and analysis that provided a view on the economics of repair and replacement for decision making based on the LCC per heat, repair costs per heat and the gunning consumption per period. The output values for application decision variables are summarized in Table 2 for each supplier's material type.

3.6 Benefit and Recommendation for Application of the Model Results

The application of the results is carried out in terms of the values determined by the model analysis for the decision criteria. Using these values of decision criteria with

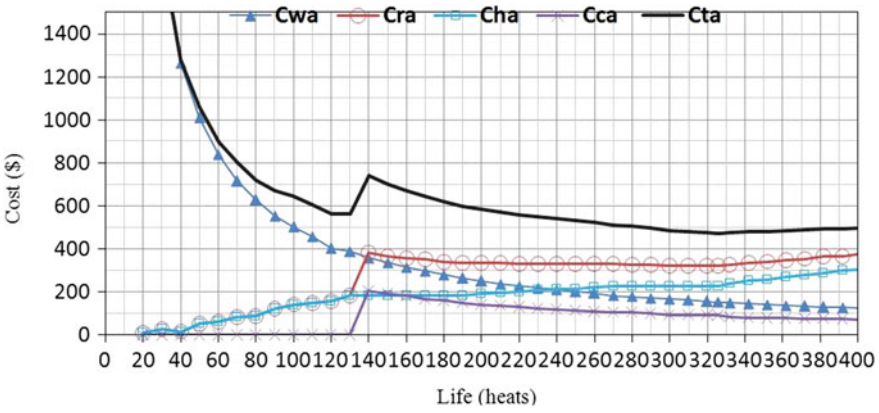


Fig. 5 Determining the optimal replacement

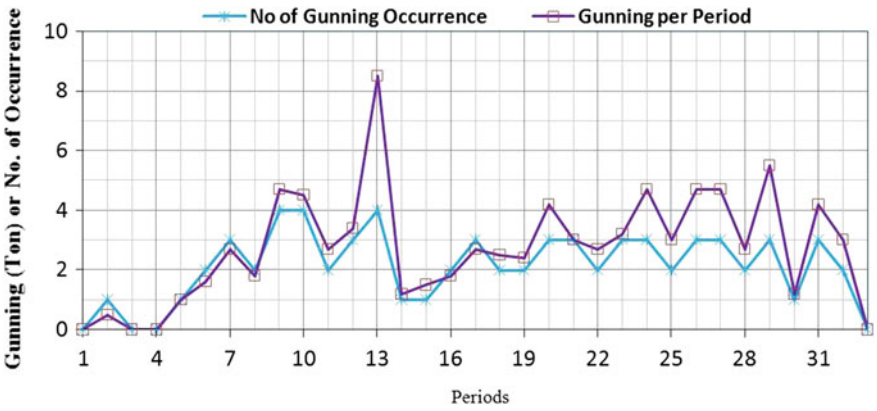


Fig. 6 Gunning consumption for hot repair sequence

Table 2 Output values of the model for decision criteria and optimum status

Parameter	Unit	Material suppliers		
		Supplier-X	Supplier-Y	Supplier-Z
Replacement cost	$	175,490	161,614	152,613
Cold repair cost	$	91,446	91,446	91,446
Maximum gunning	Ton	5.20	6.0	5.50
Hot repair period length	Heats	10	10	10
Maximum hot repair cost per period	$	13,682	14,007	15,907
Optimal EAF working lining life	Heats	278	319	229
Cold repair limit	Heats	120–130	110–120	80–90
Cold repair actual application	–	Not feasible	Feasible	Feasible
Total cost per heat (Cta)	$	1,426	1,544	1,652
Total cost per ton of liquid steel	$	15.6	16.8	16.8
Priority for use	–	First	Second	Third
Total annual cost based on use of each suppliers material alone	$	5,436,058	5,887,760	6,298,710

relation to the total management system of the company, requires fitting these criteria within the existing procedure for the decision making process at the different existing decision levels. The integration of these models is presented in Fig. 7, where all decision criteria and relevant points needed for decision making were indicated. Therefore the model is made so that, the application procedure can fit easily within the operation practice and allows for taking advantage of any production stoppage or any furnace in a non utilization case when more than one furnace exist for production.

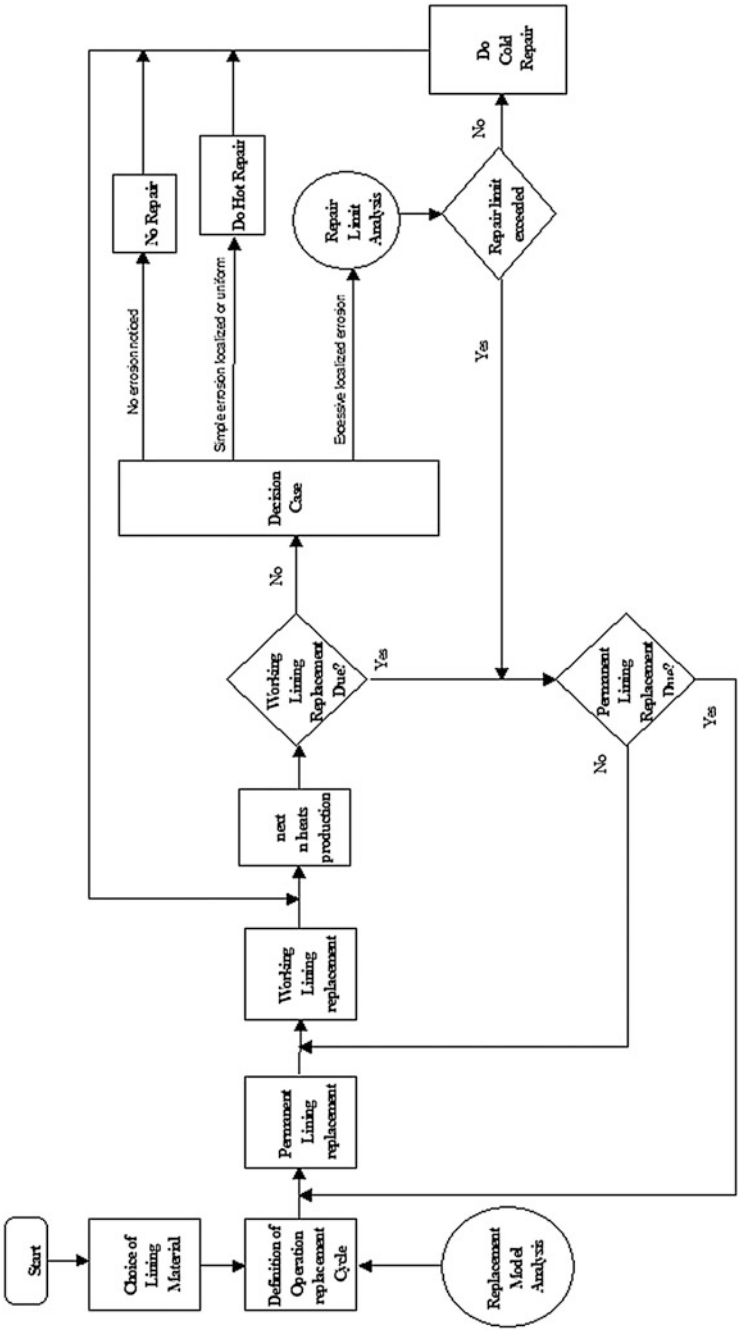


Fig. 7 Operational procedure and decision criteria

The application of the model for the optimum policy would result in cost reduction for the company:

- From 10 to 15 % annual saving in the annual refractory cost ($1,900,000) is expected when managerial decision for material type selection is based on the model output criterion.
- From 2 to 6 % annual saving in refractory cost is expected when operational decisions are based on the model output criteria.

Therefore it is an attractive proposal for top management to adopt. Furthermore the application is very simple since it only requires the use of the model findings for the economics of the decision making process with a very simple procedure that would not impose any change in the actual operation or managerial practice. The recommendation for application includes:

- Operational recommendation

  1. Always replace lining at optimum life or as close to it as possible.
  2. Observe gunning amount for every sequence of 10 heats Cold repair should not be applied if its cost exceeds the limit. Start hot repair approximately after the 20th heat for lining life less than 100 heats hot repair should be applied with a frequency ranging from every 5th to every 4th.
  3. For lining life beyond 100 heats hot repair frequency should not be less than every 3rd heat.

- Managerial recommendation

  1. When purchasing lining material, the decision of supplier selection should be based on the criterion of minimum LCC per unit production.
  2. The model should be updated in case of any future development, changing conditions or including new suppliers.

4 Conclusion

LCC is a tool to develop value-orientated decision support systems in various AM-related industries. LCC has a critical and essential role to play in asset management decision making processes.

The main results of the case study present an illustrated simple example of how LCC plays an essential role in asset management decision making; in particular replacement optimization and maintenance policy and its impact on the procurement policy. Those decision criteria in the case study are shown to be supported by the LCC-based model developed at the operational and strategic levels within the direct and support functions of the company.

Therefore, the developed LCC-based model provides a decision support system within this case study company and implies that LCC has a great potential for decision making support for asset management.

References

1. Amelsberg J (2002) Systemic performance and cost management. American Society for Quality, Denver
2. Asset Management Council (2009) Asset management. Asset Management Council and MESA Newsletter, April–May, p 4
3. Barringer P (1998) Life cycle costs and good practices. In: NPRA maintenance conference, San Antonio, Texas, 19–22 May 1998
4. Blanchard BS, Fabryky (2011) Systems engineering and analysis. Prentice Hall, Englewood Cliffs, p 585
5. Buys A, Bendewald M, Tupper K (2011) Life cycle cost analysis: is it worth the effort? ASHRAE Trans 117(1):541–548
6. Castella PS, Blanc I, Ferrer MG, Ecabert B, Wakeman M, Manson J-A, Emery D, Han SH, Hong J, Jolliet O (2009) Integrating life cycle costs and environmental impacts of composite rail car-bodies for a Korean train. Int J Life Cycle Assess. Springer, New York, pp 429–442
7. Charles AS, Alan CB (2005) Asset life cycle management: towards improving physical asset performance in the process industry. Int J Oper Prod Manag 25:566–579
8. Du Preez ND, Louw L (2008) A framework for managing the innovation process. Institution of Electrical and Engineering Computer Society, Cape Town
9. Dwight R, El-Akruti, K (2009) The role of asset management in enterprise strategy success. In: 13th annual ICOMS asset management conference. ICOMS, Sydney
10. Eginhard JM (1977) An optimal decision rule for repair versus replacement. IEEE Trans Reliab R-26(3):179–181
11. El-Akruti K (2012) The strategic role of engineering asset management in capital intensive organisations. Doctorate of Philosophy, Mechanical Engineering, University of Wollongong, Wollongong, p 247
12. Esveld C (2001) Life cycle cost analysis. In: Modern railway track, 2nd edn. MRT Productions, The Netherlands, pp 615–629
13. Farran M, Zayed T (2009) Comparative analysis of life-cycle costing for rehabilitating infrastructure systems. J Perform Constr Facil 23(5):320–326
14. Garcia FP, Lewis RW, Tobias AM, Roberts C (2008) Life cycle costs for railway condition monitoring. Transp Res Part E 44(6):1175–1187. ISSN:1366-5545
15. Ge Z, Wei W (2011) The research of comprehensive evaluation model for thermal power equipment based on life cycle cost. Syst Eng Procedia 4:68–78
16. Girsch G, Heyder R, Kumpfmuller N, Belz R (2005) Comparing the life-cycle costs of standard and head-hardened rail. Railway Gazette Int 161(9):549–551
17. Goralczyk M, Kulczycka J (2003) LCC application in the polish mining industry. Polish Academy of Sciences, Mineral Economy and Energy Research Institute, Krakow. https://www.emeraldinsight.com/journals.htm?issn=1477-7835. Accessed 16 March 2013
18. Hartman (2004) Multiple asset replacement analysis under variable utilization and stochastic demand. Eur J Oper Res 159:145–165
19. Herbert M, Gordon P (1979) Operations research techniques for management. Prentice-hall, Englewood Cliffs
20. Holland CP, Shaw DR, Kawalek P (2005) BP's multi-enterprise asset management system. Inf Softw Technol 47(15):999–1007

21. ISO/IEC Standard 15288 (2008) Systems and software engineering—system life cycle processes
22. Janz D, Westkamper E (2008) Value-oriented decision support for design to life cycle. In: LCE 2008: 15th CIRP international conference on life cycle engineering: conference proceedings. Sydney, Australia, pp 512–516
23. Jardaine AKS (1970) Operation research in maintenance. Manchester University Press, Bannes and Noble, Inc. New York
24. Jeromin I, Balzer G, Backes J, Huber R (2009) Life cycle cost analysis of transmission and distribution systems. In: PowerTech, IEEE Bucharest, 28 June 28–2 July, pp 1–6
25. Jun HK, Kim JH (2007) Life cycle cost modelling for railway vehicle. In: Proceeding of international conference on electrical machines and system, pp 1989–1994
26. Khan F (2001) Equipment reliability: a life-cycle approach. Eng Manag J 11(3):127–135
27. Kim GU, Kim KT, Lee DH, Han CH, Kim HB, Jun JT (2009) Development of a life cycle cost estimate system for structures of light rail transit infrastructure. Autom Constr 19 (2010):308–325
28. Liu H-S (2012) Analysis of LCC model of high voltage transmission system. In: Power and energy engineering conference (APPEEC), Asia Pacific, 2012
29. Mahapatra D (2008) Life cycle costs and electricity market equilibrium: a policy assessment for India. In: 5th international conference on european electricity market, 28–30 May, pp 1, 6
30. Muchiri P, Pintelon L (2007) Modelling the effects of maintenance on manufacturing performance. Int J Prod Res 45:3741–3761
31. Narayanamurthy G, Arora S (2008) An integrated maintenance and asset management system (IMAMS). Institute of Electrical and Electronics Engineers Computer Society, Bethesda
32. Norris GA (2001) Integrating life cycle cost analysis and LCA. Int J Life Cycle Assess 6 (2):118–120
33. Nyuk JW, Su FT, Raymond W, Chui LO, Angelia S (2002) Life cycle analysis of Rooftop Gardens in Singapore. Build Environ 38:499–509
34. Omar A, El-Akruti K, Afshouk A, Salama M, Barasi M, Treki S, Mansour R (1996) A company report: product unit cost of semifinished and long products
35. Patra S, Kumar S (2009) Uncertainty estimation in railway track life-cycle cost: a case study from Swedish National Rail Administration. IMechE vol 223 Part F. Rail and Rapid Transit, pp 285–293
36. Pinjala SK, Pintelon L, Vereecke A (2006) An empirical investigation on the relationship between business and maintenance strategies. Int J Prod Econ 104(1):214–229
37. Porter ME (1985) Competitive advantage: creating and sustaining superior performance. Division of Macmillan Inc, New York
38. Scarf P, Dwight R, McCusker A, Chan A (2006) Asset replacement for an urban railway using a modified two-cycle replacement model, JORS
39. Schuman CA, Brent AC (2005) Asset life cycle management: towards improving physical asset performance in the process industry. Int J Oper Prod Manag 25(6):566–579
40. Shahata K, Zayed T (2008) Simulation as a tool for life cycle cost analysis. In: Simulation conference, WSC 2008, pp 2497–2503
41. Sullivan JL, Young SB (1995) Life cycle analysis assessment. Adv Mater Process 147(2):37
42. Tähkämö L, Ylinen A, Puolakka M, Halonen L (2011) Life cycle cost analysis of three renewed street lighting installations in Finland. Int J Life Cycle Assess 17(2):154–164
43. Taylor J (2012) Asset life cycle management: case studies on asset life cycle cost modelling. Asset Management Council. http://www.amcouncil.com.au/asset-management-body-of-knowledge/asset-management-council-presentations.html
44. Thoft-Christensen P (2012) Infrastructures and life-cycle cost-benefit analysis. Struct Infrastruct Eng 8(5):507–516
45. Tsang AHC (2002) Strategic dimensions of maintenance management. J Qual Maint Eng 8 (1):7–39
46. Uppal KB (2009) Cost estimating, project performance and life cycle. AACE Int Trans 3:1–9

47. Vorarat S, Al-Hajj A (2004) Developing a model to suit life cycle costing analysis for assets in the oil and gas industry. In: SPE Asia Pacific conference on integrated modelling for asset management, 20–30 March 2004. Accessed 15 March 2013 http://ereadings.uow.edu.au.ezproxy.uow.edu.au/vorarats1.pdf
48. White J (1985) Operational research. Wiley, Chichester
49. Wijnia YC, Korn MS, de Jager SY, Herder P (2007) Long term optimization of asset replacement in energy infrastructures. In: 2006 IEEE international conference on systems, Man and Cybernetics, Taipai, Taiwan, 8–11 October 2006
50. Woodward DG (1997) Life cycle costing–theory, information acquisition and application. Int J Proj Manag 15(6):335–344 Great Britain
51. Wubbenhorst K (1986) Life cycle costing for construction projects. Long Range Plan 19 (4):87–97
52. Yu-Rong G, Chang Y, Liu Y (2009) Integrated life-cycle cost analysis and life-cycle assessment model for decision making of construction project. In: IE&EM '09. 16th international conference on industrial engineering and engineering management, pp 448–453, 21–23 Oct. Accessed 13 March 2013 http://ieeexplore.ieee.org.ezproxy.uow.edu.au/stamp/stamp.jsp?tp=&arnumber=5344554

Cost Optimisation of Maintenance in Large Organizations

Robin A. Platfoot

Abstract Large organizations routinely deploy extensive computerised systems for work management control, thereby collating databases of work order information which may be of variable quality. It is possible to analyse these data sets, taking into account issues with codification and gaps in individual work order completeness, to determine trends in work management, PM strategies, failure modes in critical plant and overall expenditure. Cost optimisation of maintenance involves three steps. The first is to understand the raw data and determine trends in what is driving the expenditure and which parts of the asset base require the most investment. The second is a codification system for the data to enable searching for possible savings across large fleets of assets. The third is to implement a search routine for savings and then apply it, to develop a credible budget savings strategy. The areas of improvement which need to be considered include work management with improvement in the planning of work so that simple or temporary fixes are replaced by well-considered and executed repairs. Secondly there is a need to improve the PM strategy so that reactive approaches to specific assets at various facilities are lifted to a more proactive approach. Both of these cases relate to addressing the issue that poor maintenance equates to high levels of work orders on assets which otherwise should not require this intensity of work resulting in wasted effort and increased cost of maintenance. In order to determine which assets have high cost levels, an internal benchmarking process can be applied to large organisations. This is implemented by comparing the maintenance of specific asset types in different facilities throughout the organisation and identifying those which have above average levels of maintenance. Results to date have shown that a credible level of overall maintenance savings may be claimed by simply identifying asset areas which are receiving too much work and then proposing improvement work to address these high rates.

Keywords Preventive maintenance · Computerised maintenance management system (CMMS) · Work management · Reliability

R.A. Platfoot (✉)
Covaris Pty Ltd, Bankstown, Australia
e-mail: r.platfoot@covaris.com.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_18

1 Introduction

This chapter is based on the proposition that increasing maintenance cost has an inverse relationship to improving asset operational reliability. In particular, reactive corrective maintenance wastes labour and resources which result in increased cost which has the potential to be managed down. There are a number of ways the operational reliability of an asset can be improved. It may be operated in a less harsh manner, incurring less damage as it is operated (e.g. back off the operation, reduce the load, and remove sources of unnecessary damage …). The preventive maintenance strategy may be improved to manage down the component of forced reactive maintenance, thereby both reducing the cost of maintenance and actually doing less work on the asset since smaller tasks are preventing larger tasks and the condition is kept to a good level. Another solution is capital change out of the asset or its major components [1].

Maintenance cost optimization will progress with the assumption that the operational demand on a facility will remain steady. If the operational requirement changes then more information is needed than can be provided from a straight budget analysis. Capital replacement should always be treated as the final option and not the first. The preference should be to conserve the organization's capital for growth purposes, and achieve more with the existing asset base [2]. There will be a limiting condition to this, and this point will be realized when reliability cannot be lifted by any maintenance-related improvement. The optimization strategy then has the following options to remove waste from the maintenance work [3]:

- Improvement of preventive maintenance strategies by specifying all of the tasks which have to be done, settings and measurements, lubrication standards and other key information. It should be noted that a balance has to be found between completeness of information provided and what the field personnel will consistently tolerate as a set of working instructions (i.e. make the document set unwieldy and it will not be used).
- Improved work management practices leading to a disciplined approach to not just return to fix broken equipment and to report on problems found which will support reliability follow-up; and
- Greater discipline in predictive maintenance and the use of condition assessments to drive high levels of condition based maintenance work, thereby reducing the level of corrective maintenance.

These three objectives should form the basis of any maintenance improvement work which can then take into account sub-projects such as, roll out of a new maintenance system, contractor reset, work management training, and condition monitoring investment. The issue is that waste is an asset-specific problem within the operating and environmental context of the asset in question. In a fleet of tens of thousands of assets the ability to isolate and classify the assets and the requirements to improve their maintenance is challenging without a specialized approach to the analysis of work order information from the maintenance system. Work order data

has the sole set of attributes that covers all of the assets, is time-stamped so that frequencies of work relate to the reliability of an asset and should carry enough attribute information to determine problems with the asset.

This chapter presents an approach to firstly classify work order data, then identify opportunities for savings and finally set up the project to resolve the issues and thereby return benefit to the organization. One of the principles of this work is the analysis accepts the quality of data provided by the maintenance teams which may be at a lower level than the system would otherwise support. Data quality is a function of what level of communication between each other the maintenance stakeholders require to do their job [4]. This is highly variable between teams and is dependent on their own use of the information such as reports, KPI trends or reliability support.

2 Work Order Information

The work order in a computerized maintenance management system (CMMS) specifies work to be undertaken on an asset and will typically document the equipment, crew to do the work, tasking, and some key dates such as when it was first raised, when it is scheduled to be undertaken and when it was closed out. Individual work orders are often highly variable in quality based on the diligence with which an individual used the system and filled out all of the details. Invariably discipline in the use of work orders and the quality of information is sustained at a high level in one of two ways:

- The work will be undertaken by third party contractors and their contract stipulates the quality of the work orders with penalties if they do not comply; and
- Work management key performance indicator (KPI) reporting tracks teams who are issuing and completing work at a standard below requirements.

To be able to use the data to identify and then quantify maintenance optimization then coding is required to track the location of the asset or assets of concern, the asset class and the type of work. The rationale is that the optimization process commits technical improvement of a specific type (based on the asset class), to assets in a specific location. The type of work is important since it will inform the nature of the improvement. There are a small number of generic work types which are possible in maintenance delivery, and they are dependent on how work was first initiated for the maintenance team. This is shown in the flow chart below (Fig. 1).

If we consider an asset which requires a high rate of maintenance work, then a small number of possibilities arise which we may consider wasteful:

- High levels of unscheduled corrective maintenance (e.g. breakdowns) means that the PM strategy is inadequate;

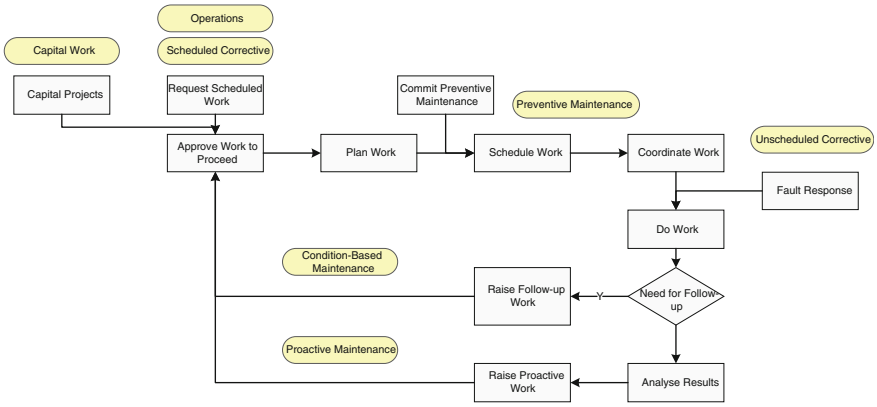


Fig. 1 Work flow and order types

- High levels of scheduled corrective maintenance means that either the asset is not fit for purpose, or more likely the people attending the maintenance are not leaving the asset in a condition which is suitable for reliable operation; and
- High levels of PMs with little in terms of condition-based maintenance being generated in response to the measured condition means that the PM schedule is wasteful and can be backed off.

Within these possibilities there are three improvement opportunities: inadequate PMs, poor work management or a wasteful PM strategy. If these issues we checked and resolved, and still the maintenance is not considered optimum, then resorting to capital improvement is a viable option. Correct codification of work type allows greater options for improving asset performance since assets under maintenance are not delivering their operational requirement. Another interesting aspect about the work flow above is the use of Proactive Maintenance, which has been borrowed from the work of R Moore [2]. In this case, if there is an absence of such work then the problem is lack of improvement using the work history and reliability engineering resources which good asset management requires.

The work order should be used to capture the costs of the job using the following information: labor hours committed by the resources which, using an average rate can be used to simply report an estimated labor cost of the work. Next, the materials required for the work should be issued to the work order within the maintenance system, so that the inventory recorded price for the goods can be allocated to the job cost. Finally special purchases of hire equipment, purchased services and specially procured materials should be allocated by linking their purchase order to the work order. It is important that every job undertaken by the maintenance teams has a reasonably accurate cost recorded in the maintenance system. This will become more evident as more of the analysis is presented further in the chapter.

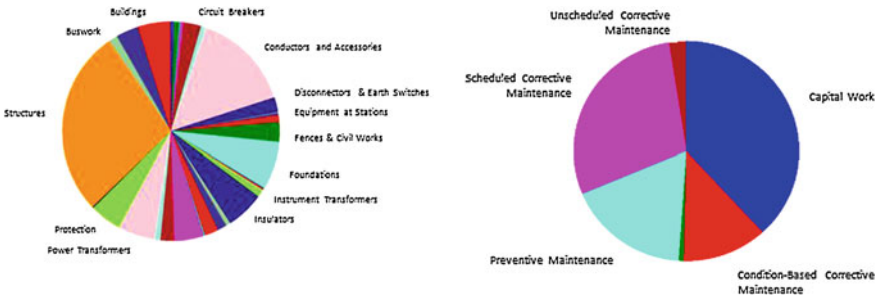


Fig. 2 Maintenance cost allocations—asset types and work types

Data exploration is always a useful initial exercise with work order data to test where the main opportunities for improvement lie. If the balances are based on cost rather than instances of work order this exercise is valuable from a commercial perspective, which will be essential in justifying any proposed improvement work. An example of cost balances by asset type for the case study material cited in this chapter is provided below with actual costs suppressed (Fig. 2).

The report on costs per asset type and work type represents an initial step in developing the strategy for where to target cost optimization. In this particular case study it was found that the bulk of the capital project work was directed to transmission structures and conductors. Savings through avoiding waste in maintenance as discussed above will focus on other types of assets noted on the left hand chart. Other considerations such as streamlining the project spend in structures and conductors on the basis of improved condition measurements are valid, but are outside the scope of this work.

3 Facilities and Internal Benchmarking

In large enterprises assets may be grouped across multiple sites or facilities. The intent is that assets are identified with a unique site and we can then test the maintenance at one site to that completed at another. The concept of the facility is shown below as a level in the asset hierarchy registered in the maintenance system. The use of the facility concept allows the analyst the ability to benchmark the maintenance within an organization, finding pockets of best performance compared for specific types of waste. If we consider the ALB facility highlighted in the example below, then there is value comparing air compressor maintenance at ALB against that undertaken at other facilities across the organization. This is a technique only suitable for large organizations where internal benchmarking of asset maintenance is used to determine standards of performance (Fig. 3).

For smaller organizations external benchmarks of so-called best practice will have to be used, and allowances for differences in trades knowledge and education,

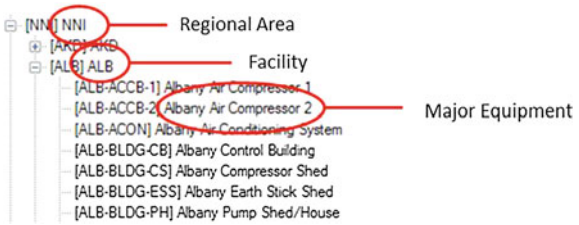


Fig. 3 Identifying facilities within the equipment hierarchy

operating environment, availability of resources and other actors all have to be taken into consideration. With the internal benchmarking concept these issues are not a consideration: if a team in one area of the organization can do well, then so can other teams working elsewhere for the same organization.

4 Work Order Statistics

A number of statistics were developed for the analysis of work order information relevant to each facility. The first measures the level of reactive work in the facility:

$$S_{WT} = \frac{(10 * PM + 7 * CBM + 3 * SCM + UCM)}{(PM + CBM + SCM + UCM)}$$

where PM is the total cost of PM work, CBM that of condition-based corrective maintenance, SCM that of scheduled corrective maintenance and UCM that of unscheduled corrective maintenance. A statistic of 10 represents a case where there is only PM work whereas a statistic close to 1 infers highly reactive maintenance.

The second statistic concerns the reliability of the assets of specific asset types within each facility, and is measured as the frequency of work which can be measured by the mean time between work orders, S_{MTBW}. If this statistic is low then the work is frequent and under the rules introduced above, a potential source of waste. The statistic is applied to all types of work: high levels of PMs may be seen as inefficient whereas high levels of corrective work are seen as indicative of less than optimum reliability.

$$S_{MTBW} = \frac{Duration\ of\ period\ (days)}{Number\ of\ work\ orders\ in\ period}$$

We need to differentiate the size of facilities before applying the reliability metric since very large facilities by the number of assets will have a lower S_{MTBW} than small facilities. In this work we found the best way to proceed was to unit rank facilities from smallest to largest in accordance with a sequence determined by a

third metric. This is a normalised statistic: the % of the number of assets at a facility divided by the number of assets at the largest facility (as measured by most number of registered assets), S_{FAC}.

$$S_{FAC} = \frac{Number\ of\ assets\ in\ the\ facility}{Number\ of\ assets\ in\ the\ largest\ facility} \times 100$$

The largest facility with S_{FAC} = 100 % is the one with the most number of registered assets. Testing for different asset types across test data showed a reasonable correlation between the mean value of S_{MTBW} with the unit rank of S_{FAC}. The unit rank means that assets are assigned a percentage between 1 and 100 % depending on where their value of S_{FAC} fell between the smallest and largest value. The 50 % unit rank S_{FAC} is the median value of the distribution of this statistic. An example is shown below for work orders associated solely with power transformer maintenance across a large number of substations in one organization. It can be seen that there is a correlation of diminishing S_{MTBW} with increasing unit rank of S_{FAC}, where the rank is greater than the median value of 50 %.

The distribution may be interpreted to mean that once a facility is large enough there is a steady load of maintenance work across multiple assets within the facility, and this load increases with larger facilities. Where the facilities are quite small (typically S_{FAC} < 10 %), then the rate of maintenance work is random and there is no correlation with the size of the facility. For the purposes of this analysis, to find significant cost savings in the maintenance across many facilities in a large organization, the small facilities need to be excised from the analysis. These may well make up 50 % of the total number but by inference they represent a smaller fraction of the total cost of the organization's maintenance due to their lower cost per unit of work (Fig. 4).

Hence this analysis should only be applied to the top 50 % of facilities organised by size across the organization. Smaller facilities will require alternative treatment.

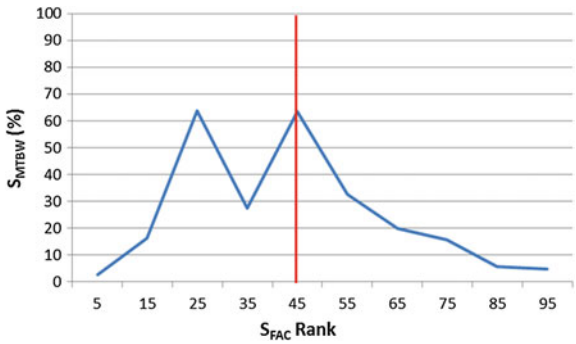


Fig. 4 Correlating Mean S_{MBTW} with S_{FAC} Rank—power transformers

5 Identifying Cost Saving Opportunities

Once the data from the work order system has been assessed for the facilities used in internal benchmarking confirmed, the statistics calculated and an appreciation of data quality determined, analysis to find cost saving opportunities can proceed. The statistical data for all of the case study organization's power transformers, distributed across multiple facilities, is presented below (Fig. 5).

Focusing on the facilities within a 80–90 % S_{FAC} unit rank, we obtain the results tabulated below. Note that facilities with a MTBW lower than the average of 5.4 days are shaded. Savings are only claimed for facilities with a lower than average MTBW, and the cost value of the saving is only calculated as if these facilities were performing at this average value and not the possible best practice value (Table 1).

The results indicate that the best performing power transformers (in terms of S_{MTBW}) located at GLN also have a relatively high level of reactive work ($S_{WT} = 4.16$). Hence the work at GLN is both reactive and infrequent. GLN is seen as a good performer at this stage because the intensity of work on the power transformers is so much lower than that at the other facilities so that its MTBW is the highest.

The availability of reasonably accurate cost data enables forecasting of savings if the work intensity at the poorer performing facilities is reduced back to the average work frequency on assets of this type at facilities of this size range. The mean time between jobs on power transformers for facilities of this size is 5.4 days. If this performance was replicated for the poor performers, then the percentage savings in actual expenditure on work can be forecast per facility as shown on the table. This results in a credible list of individual saving opportunities which can then be further investigated on a priority basis.

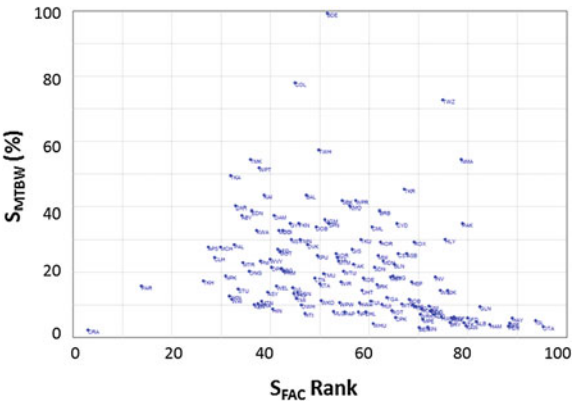


Fig. 5 Correlation of facility S_{MBTW} with S_{FAC} Rank—power transformers

Table 1 Maintenance performance—power transformers (80–90 % S_{FAC} Rank)

Facility	S_{FAC}	S_{FAC} Rank (%)	S_{MTBW} (Days)	S_{WT}	Act cost ($/year)	Revised cost (S_{MTBW} = 5.4 days)	% Saving
GLN	9.4	82.16	9.79	4.16	50,007		
HAY	14.88	88.69	6.3	8.59	64,057		
ISL	22.1	93.47	5.54	6.5	445,922		
ALB	9.25	81.41	5.15	6.42	299,239	285,385	5
BPE	14.58	88.19	4.74	7.09	128,825	113,080	12
HAM	11.4	84.17	4.31	6.4	41,071	32,781	20
PEN	13.96	87.94	3.79	6.49	291,443	204,550	30
OTA	23.55	94.97	3.64	4.81	88,757	59,829	33

6 Testing a Saving Opportunity

The investigation of an improvement opportunity should be undertaken as a consistent reliability engineering exercise. An example is provided below where the performance of transformers at a benchmarked "good" facility is compared to that of the transformers at a facility where improvement is claimed. The comparison is based on the trends in the equipment reliability and the failure modes of the corrective work. The upper plot in the set below is a Crow AMSAA of corrective work orders and the lower plot is a Pareto of the failure modes these work orders are addressing (Fig. 6).

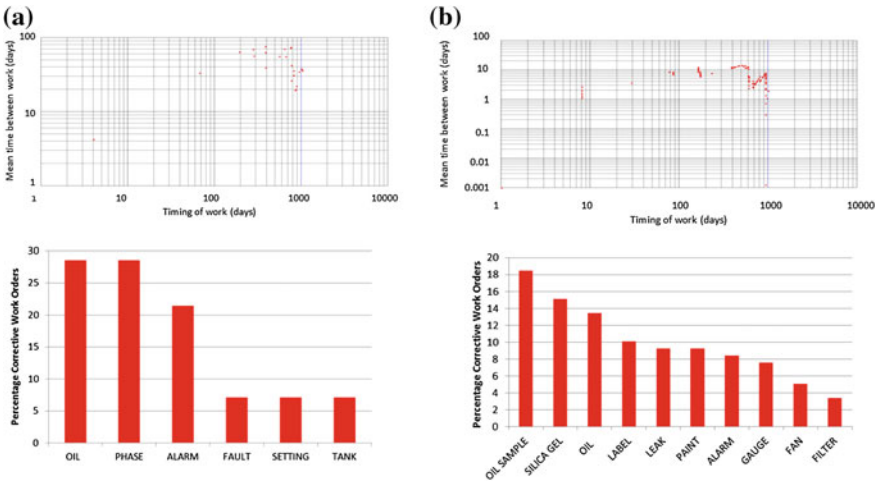


Fig. 6 Comparison of facilities' transformer maintenance. **a** HAY, **b** OTA

In the comparison we observe that the mean time between corrective work orders (MTBW) for the OTA transformers has dropped below 1 day, whereas the HAY transformers MTBW remains above 20 days between jobs. Admittedly it is of concern that in recent times the MTBW for HAY transformers has fallen and this will merit follow up in its own right. The failure modes of the two sets of transformers are particularly telling: the OTA transformers are consistently subject to oil tests using corrective work orders (rather than PMs) and the silica gel breathers are a problem area. These are simple problems which should be addressed by maintenance improvement. This comparison shows how in one case the search routine which differentiated a good performer from a poor performer was justified by the work history.

7 Recommending the Corporate Improvement Strategy

Once the algorithms have been proven for the organization's maintenance data using the logic presented in this chapter, it is then possible to automate a search for many improvement opportunities across multiple equipment types and many facilities. Considering just the case study issue of transformer alone, a summary report across multiple facilities would be presented as follows (Table 2):

There is no cost adjustment for facilities under the median value of the facility sizing (i.e. 50 % rank for S_{FAC}) owing to the issues of lack of correlation with MTBW and facility size for these small facilities. But there is a potential for significant maintenance savings associated with transformers in the larger facilities. The results above lead to a 27 % claimed savings with total reduction of costs from $ 6,877,259 per annum to $ 5,043,566 per annum. If we consider the scatter in the

Table 2 Report on transformer maintenance savings across the organization

Unit Rank S_{FAC}	No of facilities	Actual Spend ($)	Maximum MTBW (days)	Minimum MTBW (days)	Average MTBW (days)	Adjusted spend ($)
0–10%	39	406,762	2.6	2.6	2.6	406,762
10–20%	40	11,654	16.1	16.1	16.1	11,654
20–30%	40	102,096	219.2	17.7	63.6	102,096
30–40%	40	429,226	54.8	10.4	27.3	429,226
40–50%	40	867,860	1096	7.7	63.5	855,067
50–60%	39	679,476	182.7	8.2	32.5	355,312
60–70%	40	1,007,375	45.7	3.5	19.7	707,707
70–80%	40	1,963,492	73.1	3.5	15.6	940,867
80–90%	40	874,641	9.8	3.8	5.7	718,544
90–100%	39	534,679	5.5	3.6	4.6	516,332

Facility	Description	Activity	N WOs (pa)	SWT	SFAC	SMTBW	Average MTBW	Actual Cost	Saved Cost	% Savings
IGH-KIK-A	XXX	Insulators & Hardware	4	3.28	84	84	253	13669.09	4554.54	67

WO	Description	Asset	Reg Date	Cost	Type	Start Date	Finish Date	Resource	Hours
925860	MP MPE-KTA B Outage 01 May 11 Urgent Repairs Ins 315	10312063	31/05/2011	14635	IP	1/05/2011	2/05/2011	ELXNRMAKLT	
925871	MP MPE-KTA B Outage 01 May 11 Urgent Repairs Ins 519	10010891	31/05/2011	12793	IP	1/05/2011	1/05/2011	ELXNRMAKLT	
925863	MP MPE-KTA B Outage 01 May 11 Urgent Repairs Ins 510	10010853	31/05/2011	11524	IP	1/05/2011	1/05/2011	ELXNRMAKLT	
925872	MP MPE-KTA B Outage 01 May 11 Urgent Repairs Ins 550	10311787	31/05/2011	11470	IP	1/05/2011	1/05/2011	ELXNRMAKLT	

Fig. 7 Automated reporting of savings opportunity

MTBW reported in the table above, the maximum possible savings are determined where the scatter from minimum to maximum of MTBW is the highest.

This process can be repeated for all equipment types across all facilities throughout the organization. The ranking of the priorities is based on highest rate of return from maintenance expenditure to least. The identification and reporting of hundreds and possibly thousands of opportunities across the organization requires automated reporting. An example not associated with transformers but another asset type is shown in the figure above (Fig. 7).

In this example 67 % savings are claimed from corrective maintenance expenditure on the insulators and hardware on a power line circuit. The most recent work is reported along with the possible saving, which is intended to reduce a current annual expenditure of $13.7 K down to $4.55 K. The maintenance of these items is largely reactive since S_{WT} is as low as 3.28. As can be seen from the work order list in Fig. 7, the maintenance in recent time has included multiple urgent repairs coinciding during the same network outage. Instead of what looks like one annual inspection and one corrective work order leading to a MTBW of 253 days, work is being repeated on this line quarterly at around 84 days. Of the thousands of circuits throughout this particular organization, the analysis has zeroed in on this one to recommend a closer look with the maintenance team.

8 Conclusion

A registered maintenance savings opportunity refers to a specific facility, type of equipment, current expenditure to date and comparison to known average practice throughout the organization, plus the potential saving if the current performance was simply improved to the average level of performance for assets of this kind. These opportunities are ranked as amenable to one of the following savings strategies:

- Work place efficiency;
- Rationalisation of preventive maintenance procedures so that multiple work attendances are consolidated into a small efficient number of jobs;
- Reliability improvement including:
  - Change in operational procedures;
  - Optimisation of preventive maintenance to avoid failures in the future; and
  - Business case for refurbishment/replacement.

These saving opportunities have to be validated since they have been auto-extracted from the work order history which means potential misinterpretation of what actually happened in the field. The main causes of uncertainty are work management anomalies such as a high flux of jobs at about the same time, possibly which relate to one or a few causes for which multiple jobs were erroneously raised. These have to be worked through since if they are cited as a cause of the problem and not wasteful maintenance, then poor practices in control of work will be reinforced and the organization is still not efficient in delivering maintenance.

The cost savings cannot be recovered as concurrent projects no matter how well targeted. What is probably more achievable is an annual list of opportunities which can be agreed upon with the maintenance team. If this list is refreshed by analysis every year, the revision will allow for on-going refinement of the analysis process and will take into account new problems entering the organization. These opportunities should be managed through an asset management plan.

There is also a need to stage when the benefits of improvement strategies can be realised owing to the need to either establish the underlying technical capability to resolve the problems or to allow time for the benefit to show up in the maintenance current accounts. The realisation of maintenance benefits is known to follow the pattern below:

- Immediate—turn off plant which is unreliable, or consolidate work and stop repeat visits;
- Short Term—reset PM strategies to address fast acting failure modes, improve materials and component quality, or completely restore equipment rather than band-aiding problems; and
- Long Term—address work place skills and work quality, or reset PM strategies to address long term failure modes.

In closing, this analysis will not provide an automated means to actually delivering the cost savings which can potentially be found in the operations. But the analysis presented here will continue to the organisation searching across its thousands of assets to identify and then isolate a potential source of waste, which can then be targeted by traditional engineering improvement to deliver the actual saving.

References

1. Jardine AKS, Tsang AHC (2006) Maintenance, replacement and reliability–theory and applications. CRC Press, Boca Raton
2. Moore R (2013) Making common sense common practice—models for operational excellence, 4th edn. http://Reliability Web.com
3. Platfoot R, Durrant P (2008) Informed development of the maintenance work management process. WCEIAM, Beijing
4. Cappiello C, Francalanci C, Pernici B (2004) Data quality assessment from a user's perspective. IQIS 2004 Maison de la Chimie, Paris France

Multi-scale Manifold for Machinery Fault Diagnosis

Jun Wang, Qingbo He and Fanrang Kong

Abstract The wavelet transform has been widely used in the field of machinery fault diagnosis for its merit in flexible time-frequency resolution. This chapter focuses on wavelet enveloping, and proposes an enhanced envelope demodulation method, called multi-scale manifold (MSM), for machinery fault diagnosis. The MSM addresses manifold learning on the high-dimensional wavelet envelopes at multiple scales. Specifically, the proposed method is conducted by three following steps. First, the continuous wavelet transform (CWT) with complex Morlet wavelet base is introduced to obtain the non-stationary information of the measured signal in time-scale domain. Second, a scale band of interest is selected to include the fault impulse envelope information of measured signal. Third, the manifold learning algorithm is conducted on the wavelet envelopes at selected scales to extract the intrinsic manifold of fault-related impulses. The MSM combines the envelope information of measured signal at multiple scales in a nonlinear approach, and may thus preserve the factual impulses of machinery fault. The new method is especially suited for detecting the fault characteristic frequency of rotating machinery, which is verified by means of a simulation study and a case of practical gearbox fault diagnosis in this chapter.

J. Wang · Q. He (✉) · F. Kong
Department of Precision Machinery and Precision Instrumentation,
University of Science and Technology of China, Hefei 230026,
Anhui, People's Republic of China
e-mail: qbhe@ustc.edu.cn

J. Wang
e-mail: junking@mail.ustc.edu.cn

F. Kong
e-mail: kongfr@ustc.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_19

1 Introduction

It's well-known that wavelet transform is an effective time-frequency analysis tool for non-stationary signals due to its merits of flexible time-frequency resolution and efficiency of computational implementation. It has been widely applied to the field of machinery fault diagnosis because the dynamic signal of rotating machine has the non-stationary property. The signal measured under complex working conditions also contains heavy noise, which may corrupt the features extracted by wavelet transform. So some de-noising methods based on wavelet analysis have been explored [1–4]. However, those methods only aim to smooth the curves of original noisy signals, while the characteristic frequency of machinery fault cannot be directly obtained. This is because the fault-related periodic impulse is a modulator to the high natural frequencies of the machine [5]. To demodulate the impact impulses, some wavelet-based demodulation methods have been studied [6–10] and confirmed to outperform the conventional filter-based or FFT-based Hilbert transform [7, 9]. However, the extracted envelope by those methods is either at a single scale or among a scale band, which still confronts the contamination of in-band noise.

Manifold learning is an effective tool for non-linear dimensionality reduction. It has been used to extract the intrinsic manifold features of machinery dynamic systems in recent years [11, 12]. However, those features do not well consider the non-stationary property of the measured signal, which thus cannot sufficiently characterize the machinery health pattern. More recently, our group proposed the time-frequency manifold (TFM) and time-scale manifold (TSM) signatures of machinery faults, which can provide the inherent time-frequency or time-scale structure of the non-stationary signal [13, 14]. The excellent de-noising effect of TFM and TSM learning has also been verified for machinery fault feature extraction. Nevertheless, the computational load is very heavy, because each dimension of the input data of TFM or TSM learning algorithm derives from the large 2-D time-frequency or time-scale matrix. This makes these methods not applicable in on-line signal processing.

In this chapter, an enhanced wavelet-based demodulation method, called multi-scale manifold (MSM) is proposed for machinery fault diagnosis. The proposed method mainly addresses manifold learning on the high-dimensional wavelet envelopes at multiple scales. It extracts the fault-related impulses by nonlinearly combining the demodulated signals at different wavelet scales, which thus reflects the non-stationarity and non-linearity of the machinery dynamic systems simultaneously. Specifically, the MSM is produced by three following steps. First, the continuous wavelet transform (CWT) with complex Morlet wavelet base is introduced to analyse the non-stationary signal over the whole spectrum. Second, the interested scale band which contains the fault-related impulse information of measured signal is selected. Third, the manifold learning algorithm is employed to extract the envelope manifold from the wavelet envelopes at the selected multiple scales. The high-dimensional data for the MSM learning is constructed from one-

dimensional wavelet envelopes at some specific scales, which reduces the computational complexity greatly as compared to the methods of TFM and TSM. Due to its special merits mentioned above, the new MSM can exactly expose the factual envelope information of modulated measured signal; the fault characteristic frequency can be easily identified in the power spectrum of the first-dimensional data of MSM. Simulation and practical case studies confirm the effectiveness of the new method.

2 Wavelet-Based Demodulation

The continuous wavelet transform (CWT) can decompose a signal onto a time-scale plane, constructing a time-scale distribution (TSD), with each scale corresponding to a band-pass filtered signal. For a signal $x(t) \in L^2(R)$, the CWT is defined as the inner product of the signal and the wavelet functions $\varphi_{s,\tau}(t)$ as below

$$W(s,\tau) = \int_{-\infty}^{+\infty} x(t)\varphi_{s,\tau}^{*}(t)dt = \langle x(t), \varphi_{s,\tau}(t)\rangle \tag{1}$$

where $W(s, \tau)$ is the wavelet coefficient at point (s, τ) on the TSD, $\varphi^{*}(\cdot)$ stands for the complex conjugate of $\varphi(\cdot)$, and $\varphi_{s,\tau}(t)$ is generated by dilation (scale factor s) and translation (time shift factor τ) of the mother wavelet $\varphi(t)$ as

$$\varphi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\varphi\left(\frac{t-\tau}{s}\right), \quad s > 0, \tau \in R. \tag{2}$$

The mother wavelet $\varphi(t)$ is a square integrabel complex function which meets the admissibility condition

$$C_{\varphi} = \int_{-\infty}^{\infty} \frac{|\hat{\varphi}(\omega)|^2}{|\omega|} d\omega < \infty \tag{3}$$

where $\hat{\varphi}(\omega)$ is the Fourier transform of $\varphi(t)$. The complex mother wavelet $\varphi(t)$ has the property of being analytical in nature, resulting in the daughter wavelet $\varphi_{s,\tau}(t)$ being possessed of the same property. So $\varphi_{s,\tau}(t)$ can be expressed as

$$\varphi_{s,\tau}(t) = \varphi_R(t) + j\,\mathrm{H}\left[\varphi_R(t)\right] \tag{4}$$

where $\mathcal{H}[\varphi_R(t)]$ denotes the Hilbert transform defined as

$$\mathrm{H}\left[\varphi_R(t)\right] = \int_{-\infty}^{+\infty} \frac{\varphi_R(u)}{\pi(t-u)} du. \tag{5}$$

From the property of inner product and Hilbert transform, the wavelet transform in Eq (1) can be written as

$$W(s,\tau) = \langle x(t), \varphi_R(t)\rangle + j\langle x(t), \mathrm{H}\left[\varphi_R(t)\right]\rangle = \langle x(t), \varphi_R(t)\rangle + j\,\mathrm{H}\left[\langle x(t), \varphi_R(t)\rangle\right]. \tag{6}$$

Equation (6) indicates that the results of wavelet transform, $W(s, \tau)$, is also an analytical signal. Therefore, the modulus of this analytic result would provide the envelope $E_W(s, \tau)$ of the signal at scale s as

$$E_W(s,\tau) = \|W(s,\tau)\| = \sqrt{[\mathrm{Re}(W(s,\tau))]^2 + [\mathrm{Im}(W(s,\tau))]^2} \tag{7}$$

Here, $E_W(s, \tau)$ is called wavelet envelope [6]. It can be seen from Eqs. (1) and (7) that the complex CWT of a signal has the functions of band-pass filtering and envelope extraction simultaneously.

Usually, the Morlet wavelet is used as the mother wavelet for its similarity to the feature of machine fault [2]. A complex Morlet wavelet is defined by

$$\varphi(t) = \frac{1}{\sqrt{\pi f_b}} \exp\left(-\frac{t^2}{f_b}\right) \exp(i2\pi f_c t) \tag{8}$$

where f_b is the bandwidth parameter and f_c is the wavelet centre frequency. Generally, we set $f_b = 2$ and $f_c = 1$ for good resolutions in time and scale of the wavelet. A base wavelet with specific scale s is used to extract the specific frequency component f in a measured signal with the following relationship:

$$s = \frac{f_s f_c}{f} \tag{9}$$

where f_s is the sampling frequency.

As deduced above, the wavelet envelope is a band-pass filtered result at a specific scale or among a specific scale band for a modulated signal. The components outside the scale or scale band is eliminated in the wavelet envelope, however, the noise in-band can still distort the envelope of vibration signal in the diagnosis of machinery fault. To illustrate this point clearly, a simulated faulty signal from a rotating machine is constructed by considering a free vibration model with damping as follows:

$$x(t) = \sum_{k=1}^{4} A(k) \mathrm{e}^{\frac{-\zeta}{\sqrt{1-\zeta^2}}[2\pi f_0(t-k\cdot T)]^2} \sin[2\pi f_0(t-k\cdot T)] \tag{10}$$

where $f_0 = 1000$ Hz being the central frequency of the resonance band, $\zeta = 0.02$ being the damping ratio, $T = 0.02$ s being the fault repetition period, A represents initial magnitudes of the free vibration. Figure 1a shows the simulated signal with the sampling frequency $f_s = 12$ kHz. It can be seen from the waveform that four resonance pulses with different amplitudes derived from impacts at the localized fault of a machine exists periodically at the interval of $T = 0.02$ s. This kind of signal is in the form of amplitude modulation, which can be perspicuously proved by its power spectrum in Fig. 1a. where a resonance frequency band appears around the central frequency at $f_0 = 1000$ Hz with its sidebands spaced at $1/T = 50$ Hz. The scale corresponding to f_0 is calculated to be 12 by Eq (9). With this scale, the wavelet envelope is obtained through Eqs. (1) and (7) and shown in the bottom of Fig. 1a. It can be seen that the impulses related to the fault impacts are well extracted by the method of wavelet-based demodulation. However, in the presence of noise, the result may be different. Figure 1b shows the simulated signal with additive white noise resulting in a signal to noise ratio (SNR) of −10 dB. It can be seen that the noise corruption makes the periodic transient impulses in waveform difficult to be identified from the waveform. The resonance band in the power spectrum is also distorted. With the same scale as in Fig. 1a, the wavelet envelope of the noisy signal is obtained. It can be seen that although the envelope of resonance pulses in pure signal waveform in Fig. 1a can be discovered, there still exist some interfering components between the fault-related impulses, which come from the in-band noise. This may make some trouble in the detection of characteristic frequency.

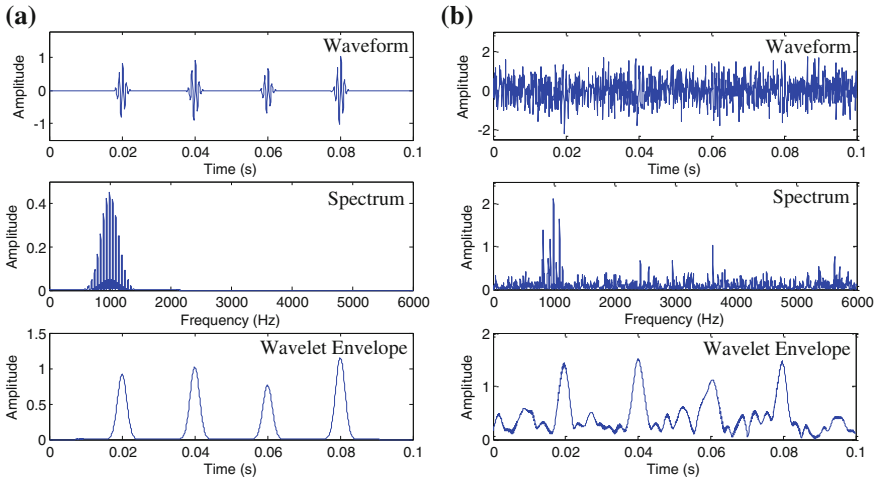


Fig. 1 The simulated signal: **a** the waveform, power spectrum and wavelet envelope at scale 12 for pure signal; **b** the waveform, power spectrum and wavelet envelope at scale 12 for noisy signal

3 Multi-scale Manifold

The problem mentioned in the previous section may be resolved by the idea of manifold learning. It can uncover the underlying structure of low-dimensional manifold from a high-dimensional data space in the process of non-linear dimensionality reduction. On account of its merits of robustness to parameters and superiority in principal manifold reconstruction, one of the manifold learning algorithm, local tangent space alignment (LTSA) [15], is introduced to reveal the intrinsic manifold feature of wavelet envelopes at multiple scales, which is called multi-scale manifold (MSM) in this chapter.

The LTSA thinks that a specific local feature of the high-dimensional data, i.e., the local projections of the neighbours on the tangent space at each data point, would be kept in the low-dimensional manifold. Thus there exists a mapping for a data set from a high-dimensional space to its local tangent space, and also from the corresponding low-dimensional manifold to the same local tangent space. For details and derivations of this algorithm, please refer to [15]. The following gives a succinct description.

Suppose D_I is wavelet envelopes at multiple scales for the dynamic signal of a faulty machine. It can be also regarded as a high-dimensional (m-D) data set with N data points as

$$D_I = [z_1, z_2, \ldots, z_N], \quad z_i \in R^m. \tag{11}$$

The LTSA algorithm would transform the data D_I to be D_O, which is a d-D ($d < m$) data set with the same data points as D_I, through the following three steps.

(1) *Local information extraction*. For each m-D data point z_i, determine a set of its k nearest neighbours including z_i as

$$Z_i = [z_{i_1}, z_{i_2}, \ldots, z_{i_k}]. \tag{12}$$

Centralize Z_i as $Z_i - \bar{z}_i e_k^T$ with $\bar{z}_i$ being the mean of Z_i and e_k being a vector of k ones, and then get the set V_i by computing the d largest right singular vectors of the centralized matrix as

$$V_i = [g_1, \ldots, g_d]. \tag{13}$$

(2) *Alignment matrix construction*. Determine a 0–1 selection matrix S_i according to the data set D_I and the neighbourhood set Z_i as

$$S_i = D_I^{-1} Z_i. \tag{14}$$

A matrix W_i is computed according to V_i as

$$W_i = I - [\frac{e_k}{\sqrt{k}}, V_i][\frac{e_k}{\sqrt{k}}, V_i]^T. \tag{15}$$

After all data points are considered by the above formulae, the alignment matrix B is constructed as

$$B = \sum_{i=1}^{N} S_i W_i W_i^T S_i^T. \tag{16}$$

(3) *Global coordinates alignment*. Compute the $d + 1$ smallest eigenvectors of B, then the d-D global coordinates D_O ($\in R^{d\times N}$) of the local tangent space is obtained with its elements corresponding to the 2nd to the $(d + 1)$th smallest eigenvalues as

$$D_O = [u_2, u_3, \cdots, u_{d+1}]^T, \quad u_i \in R^N. \tag{17}$$

The output data set D_O is a reduced version of the input data set D_I in the dimensionality. More importantly, the intrinsic manifold structure of wavelet envelope, i.e., the MSM, is exposed in D_O, especially in the first dimensional data u_2.

To extract the MSM, the wavelet envelopes at multiple scales should be firstly determined. Usually, the envelope information is embedded in the resonance frequency band of the signal's spectrum. With respect to the TSD of a signal, the envelope information in accord with the fault-related impulses distributes among a scale band on the time-scale plane. A selection method of this interested scale band is presented as follows.

Suppose $E_W(s, t)$ is the wavelet envelope matrix obtained by Eq (7) with an original scale series $S = [s_1, s_2,\ldots, s_L]$ for an analysed signal $x(t)$. The original scale band could nearly cover the whole spectrum of $x(t)$. Then the vector of the sum of matrix rows of $E_W(s, t)$ is calculated as:

$$SMR(s) = \sum_{j=1}^{N} E_W(s, t_j) \tag{18}$$

Equation (18) shows a function with scale s. s_c is the global maximum of the function, called central scale, corresponding to the central frequency of the resonance band. Then the lower limit s_l and the upper limit s_h of the interested scale band are determined as the scale value when the function value reduces to the $\sqrt{2}/2$ times of $SMR(s_c)$ on left side and right side of the central scale s_c, respectively. That is to say, $SMR^2(s_l) = SMR^2(s_h) = SMR^2(s_c)/2$. It's noted that $s_1 \leq s_l < s_c < s_h \leq s_L$. As a result, the interested scale band is selected as $S' = [s_l, \ldots, s_c, \ldots, s_h]$.

The wavelet coefficients at the selected scales form the high-dimensional wavelet envelopes, in which the MSM is embedded. The white noise is a random

distribution among the wavelet envelopes at the interested scale band, while every fault-related impulse appears along the whole selected scales band. Therefore, the impulses of fault impacts can be considered as a firm skeleton, which will be retained in the manifold learning, while the noise will be ignored in the output matrix. Therefore, the MSM can uncover the factual envelope information of machinery fault.

The main idea of the new feature extraction is to address manifold learning on the high-dimensional wavelet envelopes at the interested multiple scales, as illustrated in Fig. 2. For a vibration signal $x(t)$ of machinery with a fault, the multi-scale manifold is produced by three following steps.

First, conduct CWT on the signal $x(t)$ with the complex Morlet wavelet in Eq (8) as the basic wavelet to transform the non-stationary signal into 2-D TSD. Then we get the wavelet envelopes at all the scale band denoted as $E_W(s, t)$, $s \in S = [s_1, s_2, \ldots, s_L]$.

Second, select the interested scale band on the TSD corresponding to the resonance frequency band via the proposed method above. Then the wavelet envelopes are narrowed to be $E_W(s', t)$, $s' \in S' = [s_l, \ldots, s_c, \ldots, s_h]$. It's a high-dimensional data with the dimensionality as $m = h - l + 1$.

Third, employ the LTSA algorithm to extract the multi-scale manifold denoted as $E_M(sm_i, t)(i = 1, 2, \ldots, d)$ from the m-D wavelet envelopes. Note the dimension d is far less than m.

Since the manifold learning is a non-linear method, the MSM is the combination of wavelet envelopes at multiple scales in a non-linear way. Therefore, it can reveal the non-stationarity and non-linearity of machinery dynamic systems simultaneously. Furthermore, the computational load of MSM is largely reduced as compared to that of TFM and TSM, because the high dimensional data for manifold

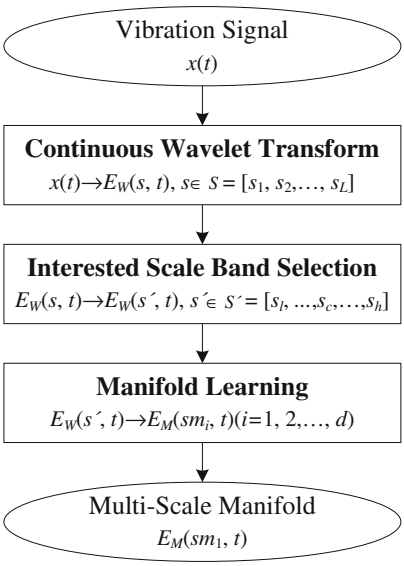


Fig. 2 Flowchart of multi-scale manifold extraction

learning in MSM is derived from only one TSD, while in TFM or TSM, multiple TFDs or TSDs are constructed for manifold learning due to phase space reconstruction [13, 14]. The structure of MSM is a d-D data, which corresponds to the wavelet envelopes at d scales. Actually, only the first dimensional data reflects the factual envelope information of machinery fault. Thus the data $E_M(sm_1, t)$ is taken for further spectrum analysis to identify the fault characteristic frequency of a machine.

To demonstrate the effectiveness of the proposed method for machinery fault diagnosis, the simulated noisy signal shown in Fig. 1b is also taken for analysis. Through the selection method of interested scale band, the TSD in the selected scale band is displayed in Fig. 3a. Each row of the TSD matrix is the wavelet envelope at the corresponding scale. It can be seen that four concentrated energies (circled by ellipses) appear periodically at the time when fault impact occurs along the whole scale band. These are the envelope information of fault-related impulses on the TSD. There also appear some other concentrated energies irregularly on the TSD, which is random information related to noise. Via the LSTA algorithm, the MSM is extracted from the wavelet envelopes at multiple scales, and the first dimensional data are shown in Fig. 3b. It can be seen distinctly that the fault-related impulses are well preserved whereas the irrelevant noise is suppressed greatly. The envelope of original signal is well revivified from the noisy signal by the proposed method, as compared to that in Fig. 1a. Therefore, the MSM is exactly suitable for extracting the actual envelope information of impact impulses of machinery fault. As seen in Fig. 3c, only one frequency component at f_d = 50 Hz and its harmonic frequencies exist in the power spectrum. The frequency f_d just corresponds to the repetition time period T of fault impacts. Therefore, it's effective to detect the fault characteristic frequency with the proposed method for machinery fault diagnosis.

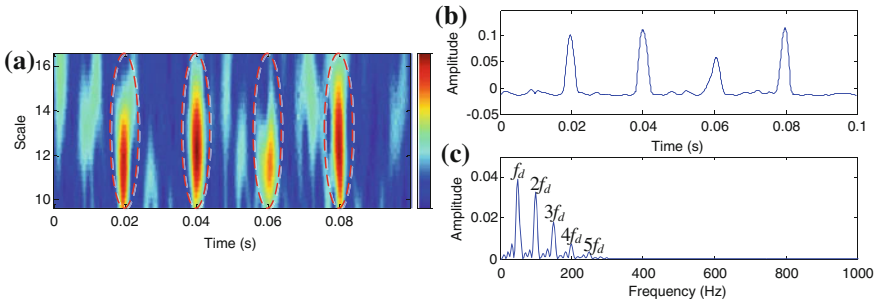


Fig. 3 The analysed results by the proposed method: **a** the TSD in the selected scale band; **b** the waveform of the first dimensional data of MSM and **c** its power spectrum

4 Experimental Verification

In this section, the proposed method is applied to analyse the vibration signal of a gearbox with severe wearing fault, to verify its effectiveness in practical machinery fault diagnosis. The experimental data was acquired from an automobile transmission gearbox, which has 5 forward speeds and one backward speed. The vibration signal is measured by using an accelerometer mounted on the outer case of the gearbox when it is loaded on the third speed, with the input rotating speed as 1,600 rpm and the sampling frequency as 3 kHz. The severe wearing faulty signal before a tooth-broken fault occurred during a fatigue test is taken for the verification of proposed method. The meshing frequency and the rotating frequency of the tested gear are calculated to be 500 Hz and 20 Hz, respectively.

As shown in Fig. 4a, the impulse feature of gearbox wearing fault is disturbed by the heavy noise in the time domain. In the frequency domain as displayed in Fig. 4b, the meshing frequency of 500 Hz shows the highest peak, but it's not demodulated. The resonance frequency band appears between 200 and 350 Hz with the rotating frequency of the tested gear as the modulator, which indicates the singularity of gearbox status. The scale band corresponding to the resonance frequency band is determined with the central scale s_c equalling 11.25 by the selection method proposed in this chapter. The TSD in this scale band is shown in Fig. 4c, where the impulse feature appears periodically, but is still contaminated by background noise. The curve of amplitudes at the black line on the TSD, i.e., the wavelet envelope at the scale s_c, is plotted in Fig. 5a, where the connection of two impulses in the waveform is discontinuous and discordant due to the noise corruption. Another popular demodulation method is also employed to analyse this faulty vibration signal. It first decomposes the signal by the method of empirical mode decomposition (EMD), then extracts the envelope information of the achieved intrinsic mode function (IMF) containing the resonance frequency components through Hilbert enveloping method. The envelope waveform of the practical signal by this method is given in Fig. 5b with the second IMF selected. It can be seen that even the impulses are distorted by the in-band noise.

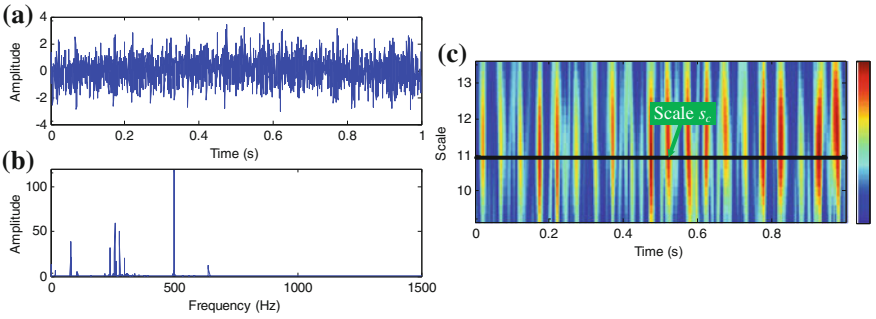


Fig. 4 The gearbox vibration signal with severe wearing fault: **a** the waveform; **b** the power spectrum and **c** the TSD in the selected scale band

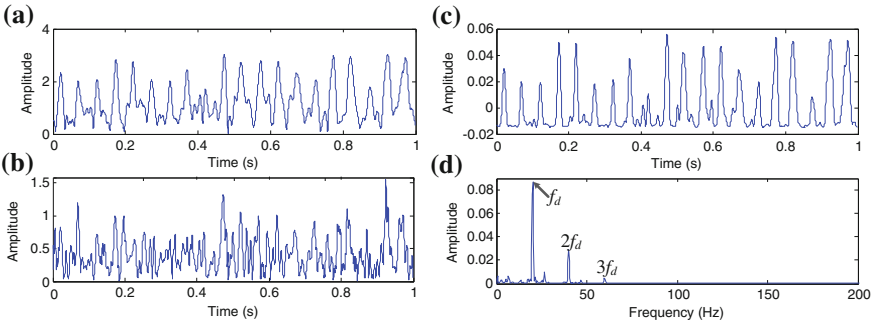


Fig. 5 The analysed results of the gearbox fault information by different methods: **a** the wavelet envelope at the scale s_c; **b** the envelope waveform by the EMD-based enveloping method (IMF = 2); **c** the waveform of the first dimensional data of MSM by the proposed method and **d** its power spectrum

To obtain the factual envelope of gearbox wearing impacts, the proposed method is introduced and the results are shown in Fig. 5c, d. As seen in the waveform in Fig. 5c, the curve is smooth and the impulses are distinguishable from each other without noise interference. The characteristic frequency f_d at 20 Hz can be easily identified through spectrum analysis since it's the dominant component of the result, as illustrated in Fig. 5d. The first dimensional data of extracted MSM captures the principal impulsive information of wavelet envelopes at multiple scales, indicating the merits in noise suppression and impulse preserving. Therefore, the MSM is more effective than the other two traditional methods used in Fig. 5a, b in envelope detection for machinery fault diagnosis.

5 Conclusion

This chapter presents an enhanced wavelet-based demodulation method called multi-scale manifold (MSM) for machinery fault diagnosis, by nonlinearly combining the envelope information at multiple scales via manifold learning. The new method concerns the non-stationarity and non-linearity of machinery dynamic systems at the same time, and can expose the intrinsic structure of envelope information of machinery fault. Moreover, the computation is efficient for MSM learning, so the proposed method can be used for longer data analysis and on-line signal processing. The first-dimensional data of the MSM matrix represents the factual envelope information of the analysed signal, showing the merits of in-band noise suppression and fault-related impulses preserving. The effectiveness and superiority to the traditional wavelet-based or EMD-based demodulation methods are verified by a study on gearbox wearing fault diagnosis.

Acknowledgment This work was supported by the National Natural Science Foundation of China (Grant No. 51005221), and the Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20103402120017).

References

1. Dl Donoho (1995) De-noising by soft-thresholding. IEEE Trans Inform Theory 41(3):613–627
2. Lin J (2000) Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. J Sound Vib 234:135–148
3. Lin J, Mj Zuo (2003) Gearbox fault diagnosis using adaptive wavelet filter. Mech Syst Sig Process 17:1259–1269
4. Qiu H, Lee J, Lin J et al (2006) Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. J Sound Vib 289:1066–1090
5. Pd Mcfadden, Jd Smith (1984) Model for the vibration produced by a single point defect in a rolling element bearing. J Sound Vib 96:69–82
6. Liu Cc, Qiu Zd (2000) A method based Morlet wavelet for extracting vibration signal envelope. In: IEEE proceedings of the fifth international conference on signal processing, vol 1, pp 337–340
7. Ct Yiakopoulos, Ia Antoniadis (2002) Wavelet based demodulation of vibration signals generated by defects in rolling element bearings. Shock Vib 9:293–306
8. Ng Nikolaou, Ia Antoniadis (2002) Demodulation of vibration signals generated by defects in rolling element bearings using complex shifted Morlet wavelets. Mech Syst Sig Process 16:677–694
9. Yt Sheen, Ck Hung (2004) Construction a wavelet-based envelope function for vibration signal analysis. Mech Syst Sig Process 18:119–126
10. Yan R, Rx Gao (2009) Multi-scale enveloping spectrogram for vibration analysis in bearing defect diagnosis. Tribol Int 42:293–302
11. Li M, Xu J, Yang J et al (2009) Multiple manifolds analysis and its application to fault diagnosis. Mech Syst Sig Process 23:2500–2509
12. Jiang Q, Jia M, Hu J et al (2009) Machinery fault diagnosis using supervised manifold learning. Mech Syst Sig Process 23:2301–2311
13. He Q, Liu Y, Long Q et al (2012) Time-frequency manifold as a signature for machine health diagnosis. IEEE Trans Instrum Meas 61(5):1218–1230
14. Wang J, He Q, Kong F (2013) Automatic fault diagnosis of rotating machines by time-scale manifold ridge analysis. Mech Syst Sig Process 40:237–256
15. Zhang Z, Zha H (2005) Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. SIAM J Sci Comput 26(1):313–338

Human Computer Interface (HCI) for Intelligent Maintenance Systems (IMS): The Role of Human and Context

Nelson Duarte Filho, Silvia Botelho, Marcos Bichet, Rafael Penna dos Santos, Greyce Schroeder, Ricardo Nagel, Danúbia Espíndola and Carlos Eduardo Pereira

Abstract This paper deals with advanced computational techniques for taking account the human factors in Intelligent Manufacture Systems. A Cyber-Physical Systems (or CPS) is a system that combines and coordinates physical and computational elements. The CPS incorporates the ability to act in the physical world with the intelligence of cyber world to add new features to real-world physical systems [1]. Among the various fields of activity of the CPS, can cite security systems, robotics, education, among others. Industrial environments are characterized by being favorable places for the introduction of technologies aimed to facilitate the interaction/mediation between human and machines. In this paper, we propose to use CPS for taking account human factors in Maintenance Estrategies.

N.D. Filho (✉) · S. Botelho · M. Bichet · R.P. dos Santos · G. Schroeder · R. Nagel · D. Espíndola
Centro de Ciências Computacionais, Universidade Federal Do Rio Grande, Rio Grande, RS, Brazil
e-mail: nelson.duartefilho@gmail.com

S. Botelho
e-mail: silviacb.botelho@gmail.com

M. Bichet
e-mail: marcosamaralrg@gmail.com

R.P. dos Santos
e-mail: rapennas@gmail.com

G. Schroeder
e-mail: sgreyce@gmail.com

R. Nagel
e-mail: ricardonagel@gmail.com

D. Espíndola
e-mail: danubiafurg@gmail.com

C.E. Pereira
Departamento de Engenharia Elétrica, Universidade Federal Do Rio Grande Do Sul, Porto Alegre, RS, Brazil
e-mail: cpereira@delet.ufrgs.br

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_20

The proposal, called TOOGLE-IMS, aims at developing a Human Computer Interface (HCI) for Intelligent Maintenance Systems (IMS).

1 Introduction

The Condition-based Maintenance approach (CBM) applies predictive techniques to avoid unforeseen breakdown of equipments. This kind of systems predict where, when and why the failure will occurs [1]. CBM systems are often used in equipments which breakdown impacts directly the production line. Intelligent Maintenance Systems (IMS) combines systems software and sensors for a CBM approach. This represents an evolution from the traditional systems of corrective and preventive maintenance to predictive systems [2, 3].

IMS new paradigm increases the role of sensor networks and computational forecasting systems that are constantly monitoring the equipments, providing support to repair and maintenance decisions [4].

On the other hand the maintenance technicians still holds many of the information associated with the operation of the equipment. The manner in which they perform repair activities may result in different future demands for maintenance. Human and context extrinsic (environmental) and intrinsic (skills, mood) factors are determining factors in activities of repair and diagnostic for future maintenance.

Thus the rescue of the role of the operator and the addition of context information to the monitoring, intelligent diagnosis and prediction of maintenance activities can contribute to improve the CBM approaches, enhancing the requirements of cost, time and quality of the processes, making IMS smarter.

In this context the challenge is to combine digital and numerical information from monitored equipments (virtual components) with qualitative and subjective perceptions from the users and the context (real components). What factors should be perceived? How to perceive them? How to integrate these real impressions to the system? These are open questions.

In recent years technological advances are leading to new relationships between humans and machines. New digitally supported environments arise allowing different agents (real and virtual) to interact in a way previously unimaginable. In this new scenario there are new theories, methodologies, techniques and tools from different areas related to Human-Computer Interface.

In this paradigm different objects equipped with embedded computing can interact each other. They can sense and adapt to the environment in a transparent manner, making the HCI simpler [5].

Researches on ubiquitous computing seek models for the interconnection of “things” (objects, computers, animals, people, etc.) in a network, similar to the devices today already interconnected through Internet [6, 7].

More recently a new approach has emerged, coining the term Cyber-Physical Systems (CPS). A CPS is a system that combines and coordinates physical and computational elements [8]. The CPS integrates the ability to act in the physical world and the intelligence from cyber (virtual) world, adding a new Human Computer Interface (HCI) and resources to real world systems [9]. Possible applications of CPS systems are in the medical, education and security areas [10].

Considering the perspective where the main maintenance tasks or fields are grouped into three primary categories: technical part, economic part and human part, our work is in the human part. Human factors associated with the maintenance management play an important role in improving the efficiency and effectiveness of maintenance processes. The operator teams have lots of knowledge about the system to be repaired which cannot be obtained solely through data analysis of sensors [11, 12].

We propose a HCI for IMS focusing on how CPS can (re)introduce the human and context factors in maintenance activities. The role of human factors in maintenance activities need to be clarified to ensure better understanding and integration of the individual knowledge and capabilities of the service personnel. This can lead to an IMS in which these information might be acquired and used for the planning of maintenance processes and forecasting the demand for spare parts.

As until now it is not possible to integrate the knowledge and the capabilities of the service personnel into the IMS, it was developed an user interface that is suitable for meeting the requirements for suchlike integration. The idea aims to explore the CPS concepts and how the human knowledge and capability can be integrated into the IMS in order to support the planning and the collaboration between the different service personnel so that the responsiveness and the skills of the IMS can be maximized.

Advanced Technologies in Virtual Reality and Sensor Networks was used to allow taking into account human factors in the maintenance activities. The proposed CPS for Human Computer Interface is based on TOOGLE platform that have four main sub-systems: i. Middleware and Components, ii. Editor, iii. Intelligent Decision and iv. Browser. We use the Editor to create and edit the different components (real and virtual) of the maintenance scenario, which is called *hyper-environment* with its components interconnected through the Middleware. The Intelligent Decision module gathers a set of applications related to the prediction maintenance. Finally the Browser allows visualization of different multi-modal information in the system.

In what follows we present a contextualization and the CBM architecture for maintenance processes (Sect. 2). Section 3 presents our system implementation to integrate both technologies (IMS and CPS). Some experiments with the proposed system (Sect. 4) and the conclusions of the paper (Sect. 5) are finally presented.

2 Condition Based Maintenance

The development of systems for Condition-based Maintenance should include the integration of a wide variety of hardware and software components.

Figure 1 shows the OSA-CBM, which provides architecture and a standard framework for implementing systems for Condition-based Maintenance [13]. This architecture has six functional blocks:

- Data Acquisition (CA) converts an output from the transducer into a digital parameter;
- Data Manipulation (DM) performs the signal analysis, calculates significant descriptors and provides the virtual sensor readings from the measurements obtained;
- Detection State (SD) provides normal "profiles", search for abnormalities whenever new data is acquired, and determines to which area of abnormality, if applicable, the data belong;
- Health Assessment (HA) provides fault diagnosis and rates of current health of the equipment or process, considering all state information;
- Prognostic Assessment (PA) determines future health states and failure modes based on the current assessment of health and use estimation of future demand for equipment and/or process, as well as the remaining useful life.
- Advice Generation (AG) provides practical information on maintaining.

Notice that the module *Human-Computer Interface* is a small module. It is proposed that maintenance routines and failure diagnosis can be improved taking in account human and context perceptions and information [14] and in this paper we intended to develop a solution for implantation of a system that can rescue and add human and context factors providing a smarter IMS.

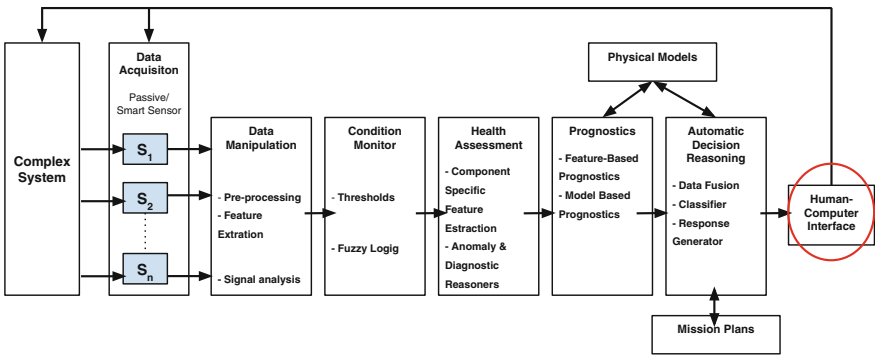


Fig. 1 OSA-CBM architecture [13]

3 Functional Requirements

In order to take into account the human factors in IMS approaches, a number of technical and social issues have to be treated. For instance, an issue is how to allocate functions among service personnel and computer systems. Another key issue is how to design a human-oriented software system to capitalize on and accommodate human skills in perception, attention, and cognition, while minimizing the opportunities for and effects of human error. Operator decision support should be designed such that the operator downloads almost all of the skill-based commands and some of the rule-based commands, while retaining the knowledge-based tasks for him.

From the IMS and HCI requirements, a set of functional requirements of our Human-Centered IMS may be identified as follows:

- Responsiveness;
- Sharing of knowledge and resources with other agents/users;
- Access to existing information;
- Dissemination of results;
- Fostering a creative environment for generate and introduce new information to the system, from the skill and mood of service personal.

The input data of the service personnel into the HCI and provided data by the HCI is the following:

- Location of the technical system (HCI);
- Product ID (HCI or service personnel);
- Bill of material (HCI);
- Damaged part(s) identified in the bill of material and/or in the visualization of HCI via e.g. a smart mobile device (service personnel);
- Required skills and equipment (HCI or service personnel);
- Description of the problem (service personnel);
- Reasons for instance due to the environment (service personnel);
- Breakdown probability or estimated breakdown date (service personnel).

An HCI-IMS may be informally defined as a "cyberspace structure of planning, control and communication mechanisms to support human decision-making via monitoring and simulation of actual maintenance situations through modeling of all activities and resources in a physical maintenance system". It can be used either as a design tool or as an operational tool. The former is used to prototype Virtual Manufacturing Systems (VMS) and can be called Editor HCI-IMS and the latter is used to simulate and control Manufacturing Systems through VMS and can be called Operational HCI-IMS.

Our approach will focus in advanced computational technologies associated with the Intelligent Environments concept. These are heterogeneous distributed sensors-actuators systems including multimedia presentation services, automation and control components, intelligent physical objects, wireless sensor network nodes,

nomadic personal or shared devices, and many other systems and entities in a Cyber and Physical System.

We introduce the intelligent environment concept through an architecture called TOOGLE, which supports the merge between various technological elements and their virtual representation, giving to designers and developers the necessary support for the creation of intelligent environments.

4 Toogle Plataform

The Toogle platform was proposed as an implementation of a technological architecture for HCI-IMS applications. It is a CPS targeted on IMS. It uses Robot Operating System (ROS) [15] for communication between real devices and the virtual world. Moreover, it uses Blender 3D rendering software [16] as object visualization tool for virtual-real industry environment. Figure 2 shows the modules that compose the system architecture.

4.1 Middleware and Components

An IMS scenario is composed by real and virtual elements that can be grouped on the following categories:

- Physical Entities: workers, equipments, tools, etc. They can be accessed by *devices* (sensors, actuators and tags); and
- Digital Entities: a set of *services* associated with logs, scheduling, digital applications, etc.

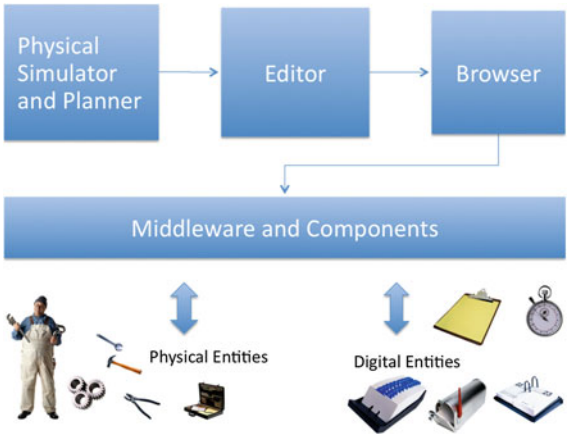


Fig. 2 Toogle platform architecture

The Toogle Components abstract the physical and virtual entities resulting in an *identifier*, a set of associated *properties* and *resources* available. Physical entities to be abstracted are called ***Smart Objects***. The Toggle Middleware is responsible for communication between the components. It is composed by the ROS system. Each *Smart Object* and *Digital Entity* is a computing process, which can run in different machines. The Middleware allows synchronous and asynchronous communication. ROS has a set of drivers for different sensors and equipments.

4.2 Toogle Editor

Toogle Editor allows building and editing cyber-physical environments, providing navigability and interaction with the information provided by *smart objects* and *digital entities*. The module allows adding and removing components (Smart Objects and Digital Entities). Moreover it is possible to edit the features of these Components, i.e., their properties, resources (devices and services) and 3D representation. For 3D representation Toogle Editor uses the Blender tool that is an open source and multiplatform system. Figure 3 show a screenshot of the Toogle Editor.

An IMS scenario is composed by a set of components and is called a hyper-environment. Digital Entities (technical reports, tutorials, datalogs, prediction, planning) and Smart Objects (workers, tools, equipments) compose an hyper-environment besides a set of Maintenance Goals (optional). Toogle uses a STRIPS-like formalism to describe proprieties, resources and goals in Hyper-environments [17].

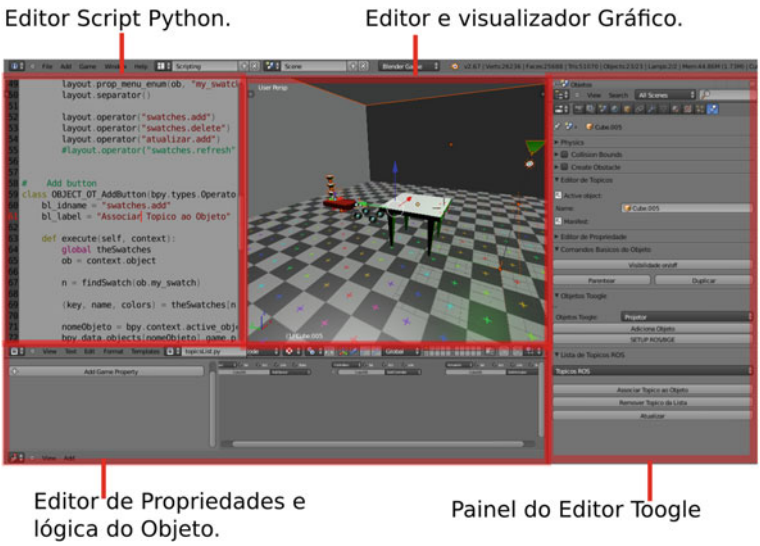


Fig. 3 Toogle editor

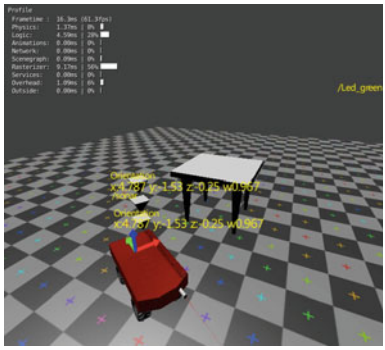


Fig. 4 Physical simulator

4.3 Toogle Intelligent Decision Module

This module provides intelligent making decision for IMS. We can add different tools and services to allow smart maintenance prediction and diagnostic.

Currently we offer two distinct services:

- Toogle Physics Simulation: it allows realism and high performance physical simulation, updating the smart object properties. It uses Bullet open source library, which uses OpenGL for real time rendering[1]; It has mechanisms for collision detection and rigid body dynamics. The physical simulator makes it possible to simulate objects that can fall, roll and collide with other objects, all with a realistic appearance. Aspects of scene lighting make use of GLSL and Pixel Shaders techniques;
- Toogle Planning: a strips planning is used to achieve the hyper-environments goals.

Figure 4 shows a screenshot of the Physical Simulator.

4.4 Toogle Browser

The Toogle browser allows the distributed and remote access and visualization of information that has been created in Toogle Editor. We can browser for Hyper-environments—their smart objects and digital entities. Toogle Browser is a multi-modal viewer with a 3D viewer scenario where a replica of a real world scenario can be seen, according to the connections created in Editor. An interface for multi projection in a CAVE visualization environment is also available. This interface allows an immersive visualization of hyper-environments. Figure 5 shows a screenshot of Browser.

[1] Bullet Physics Library. http://www.bulletphysics.com.

Fig. 5 Toogle browser

TOOGLE proposal explicitly considers the possibility of co-existence of virtual and real worlds in advanced HCI, focusing on the study of a large scalability of devices crossing/interfacing between them. The model of the architecture and implementation of TOOGLE as well as their autonomy and intelligence are being tested and validated in different scenarios involving heterogeneous devices of high scalability.

We intend to combine the novel technologies and facilities being developed to exploit and augment human knowledge and capabilities of the service personnel. The service personnel will have multi-modal, immersive 3D experiences in mixed physical-virtual worlds, including interaction with large surface displays, smart mobile devices, and wearable computers. As a result, human/context interaction with IMS will become significantly more natural, intuitive, and spontaneous than it is today.

5 Testes and Results

The platform has been tested and validated in two case studies. Real and virtual components compose a hyper-environment associated with the perception and integration of human and context factors to monitoring, diagnosis and prognosis of equipment maintenance. Next we present each experiment.

5.1 A Mobile Robot Maintenance

In this experiment we adopt as equipment to be repaired a mobile robot. This device may be considered as a complex system. The equipment is fitted with several sensors and actuators. A team of several technicians repairs the robot.

5.1.1 Components and Middleware

The experiment consists of the following components (equipment, humans and context) see Fig. 6:

- Smart objects: mobile robot, operator, toolbox and tools, tablets and computers;
- Devices: global and local cameras, GPS, RFID; and
- Digital entities: tracking, logs, and scheduler.

5.1.2 Editing Hyper-Environment

We have used the editor to create the components (*smart objects* and *digital entities*). Each component has a set of *proprieties* (position, charge level, on/off state, owner, etc.) and *resources*. *Resources* describe *devices* (cameras, GPS and RFIDs) and *services* (logs, trackers and scheduling). Strips-like predicates describe the *proprieties* and main *goals* to be achieved. Strips-like actions describe the *resources*. *Smart object* have also a 3D representation.

5.1.3 Toogle Intelligent Decision Module

We have used Bullet Physical Simulator to forecast the physical behavior of smart objects. mGTP planner [18] was used to plan a set of actions to achieve maintenance goals.

5.1.4 Toogle Browser

Figure 7 shows a snapshot of hyper-environment multi-modal representation. The browser allows accessing the different real time proprieties. For instance, location of the technical system and team, objects ID, bill of material, images from damaged

Fig. 6 Physical entities of study case: mobile robot, operators, tools, toolbox, tablets and cameras

Fig. 7 Browsing in a mobile robot maintenance hyper-environment

parts, videos and procedures from used team skills and equipment, description of the problem, maintenance forecast.

5.2 Monitoring Truckers in Shipyards

We have applied the platform in another case study related to the shipping industry. The case study was associated with the monitoring and truck repair in shipyards. We used a methodology similar to that applied in the previous example.

We decide to track trucks and workers with the use of Radio Frequency Identification—RFID. We have used Toogle Editor to create our components and their properties, resources and 3D representations. Figure 8 shows a screenshot of the hyper-environment obtained. The prototype was tested and validated, generating a database with the activities performed by all operators of the equipments and the positioning of the equipments and truckers.

In two study case real and virtual components are mixed aiming to improve the topics associated with responsiveness, sharing of knowledge and resources with other agents/users, access to existing information, dissemination of results, and fostering a creative environment for generate and introduce new information to the system.

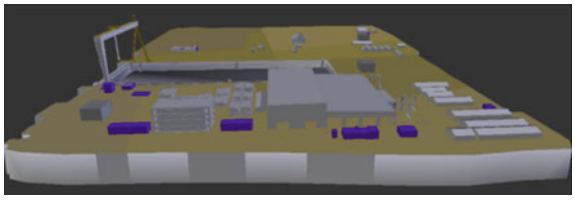


Fig. 8 Shipyard hyper-environment

6 Conclusions

This paper presents advanced technologies for HCI in Intelligent Maintenance Systems. We have proposed a CPS approach for rescuing/adding the operator/context in CBM.

We have presented the Hyper-environment concept, which is formalism for hybrid world maintenance scenario description. Toggle is a framework to design Hyper-environments. It is composed by Middleware and Components (Smart Objects, Digital Entities and ROS); Editor (Multi-modal editor, Blender, STRIP-like description); Intelligent Decision (Physical Simulator and Planner) and Browser (multiplatform, mobile, stereoscopy, multi-projection).

As future works we intend to do more tests on IMS scenarios, to treat usability and friendliness issues, and finally to improve the Intelligence of the system through a recognize system for complex events [19].

Acknowledgments The research leading to these results has received funding from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) in cooperation project, between Brazil and Germany—BRAGECRIM, named Integrating Intelligent Maintenance Systems and Spare Parts Supply Chains (I2MS2C) in the context of visualization.

References

1. Li Q, Qin W, Han B, Wang R, Sun L (2011) A case study on REST-style architecture for cyber-physical systems: Restful smart gateway. Comput Sci Inf Syst 8:1317–1329
2. Muller A, Marquez A, Iung B (2008) On the concept of e-maintenance: review and current research. Reliab Eng Syst Saf 93:1165–1187
3. You M, Li L, Meng G, Ni J (2010) Cost-effective updated sequential predictive maintenance policy for continuously monitored degrading systems. IEEE Trans Autom Sci Eng 7:257–265
4. Peysson F et al (2007) New approach to prognostic systems failures. In: Proceedings of the 17th IFAC World congress
5. Chattopadhyay D, Dasgupta R (2012) A novel comprehensive sensor model for cyber physical system: Interoperability for heterogeneous sensor. In: 2012 sixth international conference on sensing technology (ICST), pp 179–183, 18–21 Dec 2012
6. Huang Y, Li G (2010) Descriptive models for/internet of things. In: International conference on intelligent control and information processing (ICICIP), 2010, pp 483–486
7. Koshizuka N, Sakamura K (2010) Ubiquitous id: standards for ubiquitous computing and the internet of things. Pervasive Comput IEEE 9(4):98–101
8. Tan Y, Vuran MC, Goddard S, Yu Y, Song M, Ren S (2010) A concept lattice-based event model for Cyber-Physical Systems. In: Proceedings of the 1st ACM/IEEE international conference on cyber-physical systems (ICCPS'10). ACM, New York, pp 50–60
9. Yan S, Zhu Y, Zhang Q, Wang Q, Ni M, Xie G (2012) A case study of CPNS intelligence: provenance reasoning over tracing cross contamination in food supply chain. In: 2012 32nd international conference on distributed computing systems workshops, pp 330–335
10. Shi J, Wan J, Yan H, Suo H (2011) A survey of cyber-physical systems. In: 2011 international conference on wireless communications and signal processing (WCSP), pp 1, 6, 9–11 Nov 2011. doi:10.1109/WCSP.2011.6096958

11. Pophaley M, Vyas R (2013) Plant maintenance management practices in automobile industries: A retrospective and literature review. JIEM 2010 3(3):512–541. ISSN: 2013-0953; Print ISSN: 2013-8423
12. Choi B, Kim B (2000) A human-centered VMS architecture for next generation manufacturing. In : Proceedings of 2000 international CIRP design seminar, Haifa, Israel, 16–18 May 2000, pp 169–174
13. Lebold M, Thurston M (2001) Open standards for condition-based maintenance and prognostic system. In: Proceedings of MARCON 2001—fifth annual maintenance and reliability conference, Gatlinburg, USA, 2001
14. Henderson S, Feiner S (2010) Opportunistic tangible user interfaces for augmented reality. IEEE Trans Vis Comput Graph 16:4–16
15. Quigley M, Gerkey B, Conley K, Faust J, Foote T, Leibs J, Berger E, Wheeler R, Ng A (2009) ROS: an open-source Robot Operating System. ICRA workshop on open source software
16. Roosendaal T, Selleri S (eds) (2004) The official blender 2.3 guide: free 3D creation suite for modeling, animation, and rendering, vol 3. No Starch Press, San Francisco
17. Fikes RE, Nilsson NJ (1972) STRIPS: A new approach to the application of theorem proving to problem solving. Artif Intell 2(3):189–208
18. Bonet B, Geffner H (2011) mGPT: a probabilistic planner based on heuristic search. arXiv preprint arXiv: 1109.2153
19. Song Y, Kautz H, Lee R, Luo J (2013) A general framework for recognizing complex events in Markov logic. PAIR 2013: AAAI workshop on plan, activity, and intent recognition, Bellevue, WA

Using the Alliance Form for Operation and Maintenance of Privatized Infrastructures

David Mills

Abstract A significant number of privatizations utilized to operate and maintain critical networked infrastructures have failed to meet contractual expectations and the expectations of the community. The author carried out empirical research exploring four urban water systems. This research revealed that of the four forms of privatization the alliance form was particularly suited to the stewardship of an urban water system. The question then is whether these findings from urban water can be generalised to O&M of infrastructure generally. The answer is increasingly important as governments seek financial sustainability through reapplying the contestability strategy and outsource and privatise further services and activities. This chapter first examines the issues encountered with O&M privatisations. Second the findings as to the stewardship achieved by the four case study water systems are unpacked with particular focus upon the alliance form. Third the key variables which were found to have distinct causal links to the stewardship-like behaviour of the private participants in the Alliance case study are described. Fourth the variables which may be crucial to the successful application of the alliance form to the broader range of infrastructures are separated out. Fifth this chapter then sets the path for research into these crucial features of the alliance form.

1 Issues with Privatisations

Governments increasingly rely on the private sector to construct, operate, maintain and own public infrastructure. Governments had embraced New Public Management (NPM) prescriptions including privatisation which they had believed would transfer political risk and be a medicine for economic risk, public funding shortfalls, and a capability gap. Yet in Australia some of the privatisations have not been

D. Mills (✉)
Queensland University of Technology, Gardens Point Campus,
George Street, Brisbane 4000, Australia
e-mail: mills.david@outlook.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_21

successful resulting in governments bearing the political and economic risks. Governments have suffered significant political damage not only from the failures to adequately operate and maintain infrastructures such as hospitals, city train services and regional trains but also in the 'buy back' of the infrastructure where public servants and government were 'out-gunned' by the private companies. The public has not accepted that the political risk can be outsourced and has continued to hold governments accountable for critical infrastructures, triggering this research which sought to identify forms of privatization which achieve better stewardship.

An examination of the literature and archival sources revealed that the public disputation between governments and contractors over contractual scope and performance was characterised by self-interest by both parties, a short-term view by both parties, distrust by government, and an inadequate contract management capability on the part of government entities. The contractual arrangements were consistent with the NPM economic rationalist forms of relationships which focus on the self-interest of the individual organisation [5]. These led to rent-seeking and opportunism on the part of some private infrastructure operators resulting in public values such as affordability and quality of services not being satisfied. Those relationships are underpinned by traditional adversarial contracts [9] which are based on Agency theory. Agency theory is built on assumptions such as goal conflict, information asymmetry, and the agent not having an appetite for risk, yet the literature regarding emerging forms of contractual relationships [8, 10] suggested the possibility that these Agency theory assumptions are highly likely to be incorrect in the context of modern inter-organisational relationships. This research sought to identify an alternate theory which would provide a conceptualisation of the contractual relationship which does not focus on self-interest, emphasises convergence of goals and which, when applied, results in stewardship of critical infrastructures.

Stewardship theory has emerged as a counterpoint to Agency theory, emphasising a range of factors that are argued to result in the agent acting as a steward. The steward is held to act in the interests of the principal, even when the interests of the steward and the principal are not aligned [4, 6, 15]. This tenet offered the possibility that if Stewardship theory was applied to the contractual relationships underpinning privatized infrastructures then the actions and activities of the private company would be aligned with the goals of the government entity and stewardship achieved. Inherent within Stewardship Theory is the tenet that to achieve stewardship certain factors must be emphasised in the relationship. The factor *sense of responsibility* on the part of the agent offered the possibility of being highly important to the achievement of stewardship. For that reason the research took a particular focus on the operation of that sense and the actions which increase that sense.

Thus the challenge for this research was to find out:

- Whether stewardship of infrastructures was being achieved
- The behaviour that characterises a steward

- How important it is that there is a sense of responsibility to the principal for stewardship to be achieved
- The actions available to increase the steward's sense of responsibility.

2 The Findings as to Stewardship in the Case Study Privatizations

The urban water industry in Australia provided a sufficient pool of case studies i.e. two identical BOOT (Build Own Operate Transfer) contracts, one joint venture, one Alliance, and two identical management contracts which the government entity replaced with a single Next Generation (NG) alliance.

The joint venture was comprised of two equal partnerships each between the government water entity and one of two international utilities companies. One partnership distributed utilities services and the other retailed those services. The two partnerships were governed by the same board. The government entity had entered into a 20 years contract with the two joint venture partnerships for the operation and maintenance of the water system which remains owned by the government. The joint venture as steward was found to have achieved stewardship of the water system applying satisfaction of public values as the measure of stewardship, in this and all other case studies.

The BOOT (Build Own Operate Transfer) case study was comprised of two similar 25 years concession form contracts, one between the government water entity and an international utilities company for the construction, ownership and operation of a single very large water filtration plant and the other between the government entity and a global utilities and transportation company for the same functions for two smaller plants. Both contracts provided the option for the government to acquire the system by transfer at the end of the contract. Stewardship had several dimensions, the stewardship of the quality of the water, the current stewardship of the infrastructure and the stewardship at the time of the possible transfer of ownership to government. The private operators achieved stewardship of the quality of the water. There was also currently stewardship of the systems and it was found highly likely that there will be stewardship of the plants when they are eligible for transfer to the government entity.

The Alliance case study was a 10 years contract with a consortium of a private civil construction company and an engineering services company for operations and maintenance and capital works plus works for third parties in the name of the government entity. The relationship was configured in a typical modern alliance form, e.g. risk was allocated to the participant best able to bear the risk, the government entity contributed a significant number of employees, there was a *no-blame* clause in the contract, all transactions are transparent and costs and gains are shared between the parties. The alliance was found to have achieved strong stewardship of the water system yet the system remained owned by the government. This is an

important feature of this alliance model as the private ownership of water is vehemently opposed by the community [12].

The management contracts case study was comprised of two identical contracts with two unrelated private companies for O&M services in contiguous regions. Recently the government replaced these two contracts with a single NG alliance stating in the tender documentation that the management contracts had not always encouraged a co-operative focus on business improvement and that there was a lack of alignment of goals and lack of collective responsibility for outcomes. The management contracts did not achieve stewardship of the water system.

In summary the joint venture, BOOT and alliance forms of privatisation were found to have achieved stewardship of the infrastructure and the Alliance form was found to be the most effective in achieving stewardship because of the capacity to frequently (annually) adjust the performance targets and because the infrastructure could remain in public ownership if this was required by the government.

3 Key Variables that Impacted the Stewardship of the Water System by the Alliance Form

The case studies revealed features which had a causal link to stewardship of the water system. These findings not only confirm the features of the general model of the alliance catalogued by Davies [3] but highlight additional features. Collectively the findings describe a model of a contractual relationship that is highly effective for the operation and maintenance of a networked critical infrastructure i.e. an urban water system. The features were found to increase or support the sense of responsibility of the private company, a factor which the case studies revealed to be highly important, and possibly essential from the perspective of the government entity, to the stewardship-like behavior of the private participants. Accordingly the impact of the features upon sense of responsibility, and in turn stewardship will be discussed together in the following paragraphs.

3.1 Ownership and Governance

The Alliance was a *virtual* organization with each participant retaining its own assets, entering into contracts in its own name and the staff being employed by whichever participant was appropriate. The Alliance had a four member governance board of which the government entity has two members, one of which must be the chair. Contractual terms committed all parties to ensuring the highest standards of probity, transparency and open discussion of all financial and operational matters, including open-book accounting. The contract required that all decisions of the board be unanimous and *best for alliance*. These governance arrangements,

together with the mixing of employees from the three organizations were found to engender a sense of responsibility to the government entity. These features resulted in the expertise of the private sector being accessed without government surrendering ownership of the infrastructure or ceasing to maintain its own staff capability to operate the system.

3.2 Suitability of Participants

The cross-case comparison revealed that sense of responsibility of the private participants and stewardship are impacted strongly by both their own suitability and the suitability of the government entity and its staff.

All the private participants were large public companies jealous of their reputation which they protected by acting with a strong sense of responsibility to the principal. They reasoned that behaving this way would also lead to their gaining additional work with third parties by way of a business reference from the government entity. Private participants advised that their experience of other O&M contracts resulted in their knowing that they must not conduct themselves as they would for a construction contract focusing on immediate profit but rather must take a longer-term view and be co-operative and collaborative. Sense of responsibility to the government entity was inadequate in the two management contracts to such an extent that the government entity when going to market for the replacement single alliance required that one of the criterion for selection for the NG alliance was evidence of sense of responsibility to the principal in other, existing contracts.

In the Alliance government entity suitability went similarly to the issue of not taking a short-term perspective. Government entity staff took the perspective that the private companies were legitimate partners who must be commercial and make a profit. In contrast in the management contracts the government entity displayed a short-term focus upon immediate cost and entity staff were said to conduct the relationship denigrating the contractor and taking a command-and-control approach. The Alliance model obviated this command and control separation by placing a significant number of its own staff within the Alliance and requiring that the Alliance general manager and chief finance officer be appointed by the government entity.

In addition the Alliance participants not taking a short-term perspective was accompanied by strong evidence of trust, more precisely of trust which the case study informants portrayed as reciprocal and presenting in a manner not unlike the ‘high-trust spiral’ conceptualised by Fox [7]. The parties prior to entering into the Alliance contract had significant experience of each other having contracted for operations and maintenance services on a schedule of works basis, suggesting the possibility that the high trust has grown over a unusually long period.

3.3 Clear Contractual Outcomes

The cross-case comparison showed that a regime of comprehensive monthly reporting against highly specified measures engendered a strong sense of responsibility because of the clarity as to what is expected.

The Alliance was characterised by all informants asserting that annual adjustment of the targets for performance indicators and the incentive payment being tied to the achievement of these targets led to a highly focused sense of responsibility. The joint venture targets were adjusted 4 or 5 yearly by negotiation with the government entity, this period aligning with the re-set of the prices by the independent regulator. The joint venture advised that the inability to adjust the targets more frequently was onerous because of the frequent increase in costs, e.g. energy, chemicals and labour. The Alliance annual adjustment offers greater opportunity for all parties to address risks.

3.4 Incentives to Private Participants

Incentives had the greatest impact on sense of responsibility, confirming the Agency Theory model of agents (the private companies) being motivated to act in the interests of the principal by extrinsic rewards.

Financial incentives in the BOOTs case study were in the form of a usage charge which varied with the quantity and quality of raw water delivered to the plant operator or delivered by the operator. This charge nicely fits the risk to the operator if the raw water is of poor quality and penalises the operator if it supplies water which is below the quality required in the contract. The informants asserted that this regime led to the companies acting with a sense of responsibility. In the Alliance the capital work was incentivised by a gain share/pain share payment. For O&M work the margin added to all work performed by the companies was increased if the annual targets are met.

A longer-term contract was very important to private participants and was seen by government entity informants as an incentive to act with a sense of responsibility. The terms of the contracts in the case studies where stewardship of the water systems were achieved were 10, 20, and 25 years. The informants believed these contracts were attractive to private company owners when compared with construction contracts which were typically 2–3 years. The longer term of water O&M contracts allowed the companies to establish their operations and in subsequent years harvest significant profit through efficiencies gained by investment in technologies, negotiation of reduced prices for inputs. The contracts typically had provision for extension which was seen as an incentive for acting with a sense of responsibility and for high quality performance. The longer term of the contract meant the governments were able to hold the private companies accountable for work carried out earlier in the contract term.

Additional works, either provided for in the contract or gained from like water systems by association with the government entity was found to be a powerful incentive to act with a sense of responsibility. The Alliance contract provided for additional work such as capital works at the discretion of the principal. The Alliance contract also allowed the Alliance to provide services to third parties such as smaller water systems in the name of the government entity. This work has been grown to a significant value and has allowed the civil engineering private participant to accelerate the industry uptake of its propriety technology. The Alliance private participants advised that their companies adopted the strategy of acting strongly in the interests of the government entity in the expectation that a favorable reference from the government entity would lead to work from like water systems.

4 The Possibility of Generalisability to Other Infrastructures

These features of the Alliance were found to make the alliance form of privatisation highly effective in stewardship of the urban water infrastructure whilst allowing the government to retain ownership of the system. However, the issue of infrastructure ownership and many of those features are either typical of urban water, e.g. the ease of specifying the outcomes and the scrutiny by independent external bodies, or unique to the particular case study, e.g. the parties having established reciprocal trust pre-contract. The task then is to distinguish the features which might limit the generalizability of the findings to other networked infrastructures or other infrastructures which are not core to the government entity's purpose but nonetheless are essential. The features which contributed to the stewardship by the alliance form will first be drawn together in a taxonomy and then those which have the possibility of not being present in other candidate privatisations will be discussed.

Features of alliance privatisation which contributed to sense of responsibility and stewardship:

- Risk being allocated to the participant best able to bear that risk
- Contractual outcomes specified with great clarity
- Transparency of transactions between alliance participants
- Information symmetry
- Intense measurement, rigid reporting and incentives
- Transparency to the public through external reporting and independent scrutiny
- Long term of contracts
- Contributing a significant proportion of the alliance staffing
- Culture and capability to establish and manage contracts with private sector
- Reciprocal trust.

These features fall into two categories, namely those features of a candidate privatisation which are readily assessed and those that are difficult to assess but their absence in the candidate may be critical to the possibility the privatisation

succeeding. An example of those which are easily assessed is the extent to which outcomes can be clearly specified and internal and external reporting instituted. Those that are not readily assessed are the culture and capability of the government entity to establish and manage the contract, and the capability of the government entity staff to engender reciprocal trust with the private company staff. Because the literature and understanding of government entity culture and capability, and the operation of trust in the context of alliance privatisations is underdeveloped a significant risk to the success of privatisations is presented.

The government entity having the culture appropriate to participate in an alliance was demonstrated by the Alliance case study. The management contracts case study demonstrated that both the governance arrangements and the perceptions and conduct of individual government entity staff who interfaced into the contracting companies could negatively impact the success of the contract. A second, key aspect of the culture that was inappropriate was the perception by government entity staff that contractors were not to be trusted and that a command-and-control approach was appropriate, rather than the collaboration found in the Alliance case study. The government entity recognised these deficiencies and replaced the management contracts with a single alliance with a single governance body and configured the NG alliance so that 30 % of the alliance staff were from the government entity. This finding of inappropriate perception as to the alliance relationship suggests that for government entities there may be a continuum of suitability for participation in privatisation, ranging from the highly effective approach taken in the Alliance case study down to a combative, normative Agency Theory model characterised by an absence of trust.

Trust was palpable in the Alliance case study which had the benefit of the three parties to the Alliance contract having earlier experience of each other in the schedule of rates era which preceded the Alliance contract. Trust has been found to mitigate the problems in privatisations that arise from lack of competition and difficulties in specification and monitoring [1]. However Alford [1] found that turbulence (frequent change in government organisations), complexity across government, accountability requirements upon government entities and staff, and the differing organisational cultures between government and private partners presented as obstacles to building trust. Alford [1] observed that as these factors were inherent to public management it is likely that they could not be eliminated, but rather only ameliorated. Alford [1] suggests that the amelioration could be achieved by: selection of staff who have particular skills e.g. networking, negotiation and influencing and values which do not jar with those of the partnership and by changing the structure of the government organization so that it better fits with the partnership. Alford [1] provided examples of how those structural and people changes might be made but much remains unanswered as to how to determine where on the suitability continuum the government entity currently sits and where it must be before commencing the privatisation of the activity.

The capability of the government entity to establish a contract on the appropriate terms and in turn manage that contract so that the required outcomes are achieved was displayed in the Alliance, joint venture and BOOTs case studies. The

government entity in the BOOT's achieved strong contract management allocating the responsibility to the executive position responsible for the delivery of the outcome, i.e. safe drinking water, and adequately resourcing routines that involved the several levels of interfaces between the organizations. In the joint venture the 50 % ownership by the government entity and contractual requirements as to comprehensive reporting and information symmetry allowed the government entity to manage the contract very effectively using limited resources. In the Alliance, features such as staff embedded in the Alliance workforce, the contract specifying that the Alliance general manager and chief finance officer are to be chosen by the government entity only and the government entity retaining outside the Alliance the capability to plan and design the system resulted in strong contract management.

This capability to organise the management of substantial contracts with private providers is typical of the urban water industry and critical networked infrastructures generally. Yet privatisation is increasingly being applied to infrastructures and services which support, but are not themselves, the core activity of the government entity [14]. Examples are the buildings and pathology and housekeeping services required for a modern hospital or the provision of information technology to community services agencies. Contract administration has been described as the *neglected stepchild* [11] and governments have been found to be reluctant to invest in the capacity to manage outsourced services [2]. For that reason it would seem desirable that a better understanding of the components of appropriate contract management be identified with sufficient specificity to establish a framework to be applied to candidate privatisations.

5 The Research Path

That there was strong capability within water entities to establish and then manage contracts for privatised services is unremarkable in that managing such systems and putting to contract substantial construction projects is what those entities have done for hundreds of years. What is remarkable is the expectation of governments that there will be no significant political risk from entities which have limited or no experience establishing and managing contracts with massive international companies which are deeply experienced in managing contracts with governments. To assist in minimizing the number of privatisations which fail because features of the government entity are inappropriate this chapter will establish the path of research to further develop the knowledge of the alliance model of privatisation from the perspective of government.

The areas of research are:

- The capability of government entities to establish and manage an appropriate long-term contract.
- The culture within the government entity including the capability and capacity to establish and maintain a high trust environment that is suitable to participate in an alliance relationship.

The methodology for this proposed research will not be fixed until completion of the first stage, an exhaustive exploration of both the immediate literature, e.g. privatisation, culture [13] or reciprocal trust [7] and the broader, related literature extending into areas such as procurement, culture and change management.

In respect of the capability to establish and manage an appropriate contract it is proposed that the objective of the first stage will be to assemble a framework comprised of the component elements and skills involved in specifying the service to be provided by an alliance and then creating and conducting an effective tender process. A second, similar framework of functions and skills required to effectively manage an ongoing contract will be developed during this stage. Dependent upon the extent of the information available from this exploration of the broader literature it may be possible to form both these frameworks. If that is possible then a logical next stage would be the testing of these frameworks by way of application to case study privatisations.

The dimensions of the culture required to give effect to that management of the alliance relationship will be developed utilising the hierarchy of organisational culture (the tangible artefacts, the professed values and the tacit and unspoken values) established by Schein [13]. The specific dimension, the preparedness of government entity staff to trust private service providers, will be explored commencing with the work of Fox [7] and Alford [1]. The objective is to establish the dimensions of that trust and provide an accompanying framework of factors which impact trust.

Accordingly this research points scholars, government and industry towards the use of the alliance form of privatisation for a broad range of infrastructure privatisations. This chapter sets the path for further research aimed at increasing the knowledge of the capability of government entities to establish and manage contracts and the dimensions of the culture which supports the appropriate management of alliance privatisation contracts.

Acknowledgements This project is supported by the Cooperative Research Centre for Integrated Engineering Asset Management (CIEAM).

References

1. Alford J(2009) Tackling inherently governmental obstacles to building inter-organisational trust. In: XIII annual conference of the international research society for public management, Copenhagen, 6–8 April
2. Brown T, Potoski M (2003) Contract-management capacity in municipal and county governments. Public Adm Rev 63(2):153–164
3. Davies JP (2008) Alliance contracts and public sector governance. Doctoral thesis Griffith University, Queensland. http://www4.gu.edu.au.80/80/
4. Davis J, Schoorman F, Donaldson L (1997) Toward a stewardship theory of management. Acad Manag Rev 22(1):20–47
5. Denhardt RB, Denhardt JV (2007) The New public service: serving, not steering. M. E. Sharpe, Armonk

6. Donaldson L, Davis JH (1991) Stewardship theory or agency theory: CEO governance and shareholder returns. Aust J Manag 16(1):49–65
7. Fox A (1974) Beyond contract: work, power and trust relations. Faber, London
8. Grimshaw D, Vincent S, Willmott H (2002) Going privately: partnership and outsourcing in UK public services. Public Adm 80(3):475–502
9. Keast R, Waterhouse J, Brown K, Mandell M (2005) Hard hats and soft hearts: relationships and contracts in construction and human services. In: EGPA conference, Berne, Switzerland, August 2005
10. Keast R, Mandell M, Brown K (2006) Mixing state, market and network governance modes: the role of government in "crowded" policy domains. Int J Org Theor Behav 9(1):27–50
11. Kelman S (2002) Contracting. In: Salamon LM (ed) The tools of government: a guide to the new governance. Oxford University Press, New York
12. Prasad N (2006) Privatisation results: private sector participation in water services after 15 years. Dev Policy Rev 24(6):669–692
13. Schein E (1985) Organizational culture and leadership. Jossey-Bass, San Francisco
14. Sturgess G (1996) Virtual government: what will remain inside the public sector? Aust J Public Adm 55(3):59–73
15. Van Slyke DM (2007) Agents or stewards: using theory to understand the government-nonprofit social service contracting relationship. J Public Adm Res Theor 17:157–187

A Resources Provision Policy for Multi-unit Maintenance Program

Winda Nur Cahyo, Khaled El-Akruti, Richard Dwight and Tieling Zhang

Abstract A model for resource provision policy in multi-unit maintenance program is developed. The model is based on integrated System Dynamics modelling methodology. First, a scheme is developed to formulate and evaluate policies regarding the provision under uncertain condition or resource requirement in a period of planning to result in minimum maintenance cost. The scheme then is transformed into the suggested research model. In the modelling process, emphasis is placed on resourcing strategy for maintenance involving human and procurement of spares and any other materials. The model is composed of three sub-models that interact based on causal loop modelling between variables. The human resource sub-model focuses on policies related to provision and management. The procurement sub-model takes into account all critical spare parts, tools and equipments, and any other supporting material. The interaction between variables with sub-models is then evolved into a multi-unit maintenance resource provision model. The developed system dynamics model has been verified using data from a case study. Due to data availability, the verification has encloses only the relation between maintenance and human resources but further application on wide range of resources through case studies is intended. The advantage of this model is related to its wide outlook of interrelated variables of a complex multi-unit maintenance system. This model can show the impact of one decision in a certain maintenance

W.N. Cahyo (✉) · K. El-Akruti · R. Dwight · T. Zhang
University of Wollongong, Wollongong, NSW 2522, Australia
e-mail: winda.nurcahyo@uii.ac.id

K. El-Akruti
e-mail: khaled@uow.edu.au

R. Dwight
e-mail: radwight@uow.edu.au

T. Zhang
e-mail: tieling@uow.edu.au

W.N. Cahyo
Department of Industrial Engineering Faculty of Industrial Technology,
Islamic University of Indonesia, Yogyakarta, Indonesia

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_22

resourcing system on the other maintenance resourcing system and will lead to an optimum policy for maintenance resource provision system.

Keywords Resource provision · Multi-unit maintenance system · System dynamic modelling

1 Introduction

In engineering asset management, maintenance has become a mainstream of research focus. Most research is focusing on development of more effective, efficient and optimum maintenance systems that will contribute to better performance of assets. A broad range of models to enhance and achieve optimization in maintenance practice has been developed e.g. [1–4]. Maintenance resources play an essential role in such optimization. Optimization of maintenance practice becomes difficult for complex multi unit maintenance system. This complexity involves ensuring adequate maintenance resources and sufficient allocation of the maintenance resource to each unit to cover the requirement of maintenance process and guarantee that all units are able to achieve the desired reliability.

This research aims to model the process of managing maintenance resources in a multi-unit maintenance program for developing an effective maintenance resource policy. In multi-unit technical system, resource allocation is integrated into maintenance programs by synthesizing each unit along with the required resources from the maintenance program. In this manner, all required resources are accumulated into the total resources required for the whole technical system as a part of an integrated maintenance planning program. The required amount of resources as a result of maintenance resource planning has to be compared with the available maintenance resource. This process is similar to the aggregate planning process in manufacturing industry. It is a common situation that the number of maintenance resources becomes a limiting condition for a technical system to achieve a certain performance that must fulfil the business needs. In order to make a policy for such situation, an appropriate modelling approach is required.

In general, a large number of publications on maintenance method and resources exist but the nature of each industrial system requires a unique maintenance resource management system in terms of type, capacity and complexity [1]. This argument leads to the need for undertaking research in developing a specific model for the compatibility of maintenance resource management with the particular nature of the industrial technical systems. In a multi unit technical system, each unit may require different maintenance policy that involves different amount of maintenance resources over time. The implementation of such maintenance policy in one unit will affect the availability of maintenance resources for the others. Unavailability of required maintenance resources may lead to ineffective maintenance programs and may cause unit failure [2].

A range of research models have been developed for multi unit maintenance system, either utilizing analytical solution or simulation. Models used for multi unit maintenance system that utilize analytical solution are usually referred to as mathematical models [3–5]. In a certain complexity of a technical system, a mathematical model can be adequate to model the system, however; when the system is getting more complex, the use of simulation is a preferable option [6]. Altiok and Melamed [6] argued that it is difficult to find a proper model to represent the system being observed, and developing such model can be expensive. Endrenyi [7] stated that the mathematical modelling involves a large number of input information that sometimes is not easily obtained. For those reasons, the decision makers will be likely to avoid complex mathematics and modelling techniques that time-based data [8] and consider simulation as an alternative for modelling.

Beside the system complexity of resource allocation in a multi unit maintenance system, the time horizon of the policy must be considered in the modelling process because it is a dynamic system where the maintenance resources' states always change over time [9, 10]. This situation requires a detailed analysis of requirement, provision, and allocation of maintenance resources and therefore a systematic and dynamic maintenance resources policy model is required. The modelling technique must be able to capture the dynamics of the system to describe the effect and feedback of the maintenance policy of each unit to the overall technical system control [10]. Based on this analysis, system dynamics methodology tends to be an appropriate method to deal with the development of a model for the maintenance resource policy making purpose.

2 Research Outlook

The implementation of any maintenance resource policy for a unit in a system will influence the other units directly or indirectly. According to Yang et al. [11], for making a good decision regarding to this maintenance decision making, it is important to have a good structure of the complex technical system to allow for analysing the important relationship among elements in the system and sub-system. The system structure must *also allow for* explaining the feed-back or consequences of a certain implemented decision to the whole system performance.

It is proposed that system dynamics can be useful for enhancing the maintenance infrastructure to improve maintenance performance. Kothari [12] and Xiaohu et al. [10] developed system dynamics models for maintenance system analysis. Kothari [12] developed a system dynamics model in order to analyze the dynamics behaviour of maintenance policy's components in the area of system's economic and technical performance. He defined the dynamics of behaviour and then formulated and tested alternative policies to improve system performance. Similarly, Xiaohu et al. [10] also developed a system dynamics model to analyze maintenance system's basic elements and structure for multi components technical system. Both Xiaohu et al. [10] and Kothari [12] in their model have apparently assumed the

number of maintenance resources is unlimited. More advance, Bivona and Montemaggiore [13] developed a model that connected maintenance system with human resource, finance, service provision and assets management systems. The purpose of the modelling is to observe short and long term implication of such maintenance policies in a city bus company. The model included human resources management system to cover maintenance process however did not include other maintenance resources which made it impractical for industrial multi unit maintenance system. Considering the importance for managers to understand the dynamics behaviour of maintenance system [12] and see its practical implications [13], this chapter focus on developing a system dynamic model to support managers deal with the resource provision policy making for a multi unit maintenance system.

3 Development of the Model

From system modelling perspective, the model of maintenance system policy making purpose consists of input, process and output. The input of maintenance system can be in the form of maintenance system information of all its resources such as human resource, spare part, equipments, and machine information. The maintenance process involves interactions between the elements: machines/units, human resources, parts, and tools and equipments. The output is related to the purpose of the policy analysis.

In this research, the main purpose is how to achieve the desired system performance measured by maintenance system reliability while minimizing the cost of maintenance through an appropriate resources provision policy. Representation of the analysis for the resources provision policy in maintenance system is presented in Fig. 1. As seen in Fig. 1, to make a policy for maintenance resource provision, information about the state of overall maintenance resources and the desired overall performance is required. Based on this information, a set of alternative policies can be made and compared. Generally, the selected policy can be implemented either in terms of recruitment or procurement policies, or in terms of process adjustment policies that impact the maintenance system output or the desired maintenance system performance.

The main objective of this research is developing a system dynamic model that can be used to analyze the maintenance system in order to make a suitable maintenance resource provision policy. The model should examine the current approach of maintenance resources provision and develop scenario for better resource provision policies for comparison. The performance of the current resource provision policy is compared with the suggested scenario developed by the system dynamic model to find the best for implementation.

As discussed, system dynamics methodology is used to model the resources provision policy for the maintenance system. The system dynamics methodology consist of five phases as presented by Maani and Cavana [14].

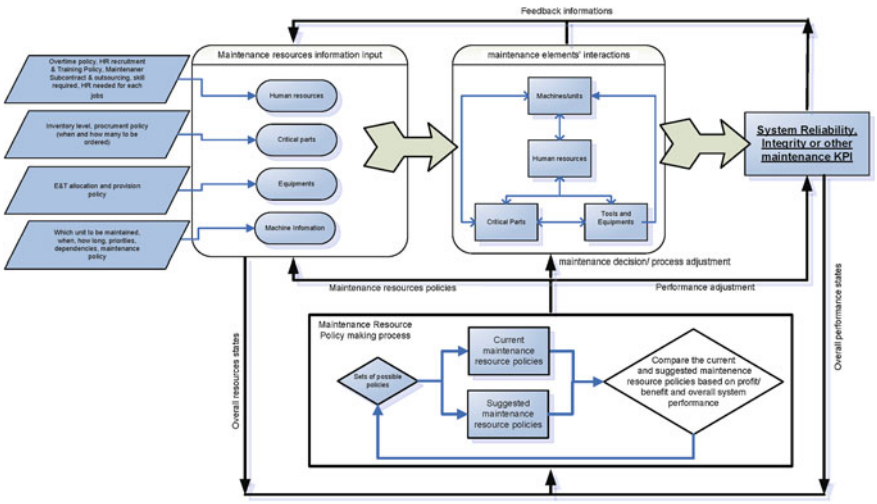


Fig. 1 Representation of analysis for maintenance resources provision policy

In the early modelling process in Maani and Cavana [14] causal loop modelling or causal loop diagram (CLD) is presented as a conceptual model. In this chapter, the CLD is developed based on the same concept associated with maintenance system modelling. The focus of this chapter is related to human resource management system and procurement system. The human resource management system caters for a policy on human resource provision which manage man hours based on availability and assignment. The procurement system deals with purchasing of spare parts, consumable materials, tools, and equipments or contracting with service providers. A simple dynamic model of human resource management system, maintenance and procurement system is as shown in Fig. 2.

The model is built based on the observation of unit failure rate in relation to the causal process of preventive and corrective maintenance and the maintenance resource requirement of each maintenance process in each unit. Failure rate increases overtime and can be reduced by preventive and corrective maintenance (loop B1 and B4). Preventive maintenance is performed based on the schedule (PM schedule) with the required man hours while man hours for corrective maintenance are allocated as required. Man hour required to perform maintenance task over time are compared with the available man hours. The result of the comparison is a number of man hours to be assigned to each maintenance job (Loop R3 and R4).

In human resource management system, available man hours are reduced by absence or leave but available man hours can be increased by overtime, outsourcing and new hiring.

In procurement system, availability of resources is influenced by the requirement of those resources for maintenance purposes and the quantity and number of resources' orders launched to suppliers. The number of order quantity is considered based on qualitative parameter (e.g. expected demand, desired inventory level,

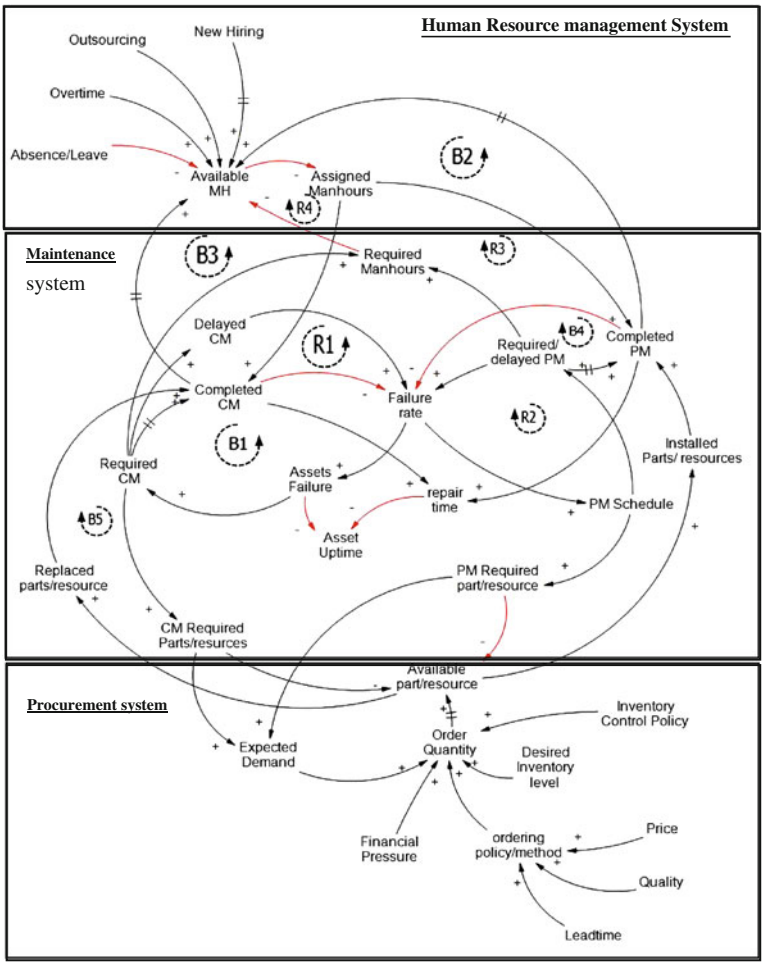


Fig. 2 CLD for maintenance and related system

Inventory policy control) and qualitative parameter (e.g. financial pressure, ordering policy/method). Expected demand is compiled from projection of maintenance activities in the future based on historical data owned by management.

For illustration, the modelling technique is mapped for one unit technical system as shown Fig. 2. The general model for multi unit maintenance system is presented in Fig. 3 where each system has its strategy to achieve the desired system objective. The strategies are implemented in each system ensure that optimal conditions are achieved. In a system approach, the system is not considered as the sum of elements but the optimum integration of each element to attain the system objective effectively and efficiently. In this perspective and as reflected in Fig. 3, the best system performance may not be the accumulation of each best strategy in each system but the best combined strategy. To determined the best combined strategy for overall

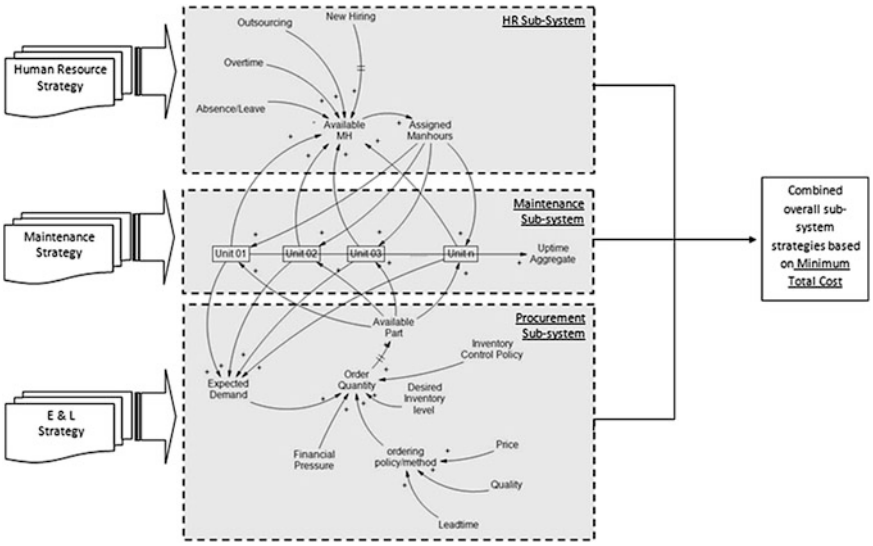


Fig. 3 Relation of multi unit maintenance system with other systems

systems, each strategy in each system must be simulated in the model to know which combination has the best outcome related to the objective of the policy making process.

The idea presented in Fig. 3 is very complex. The complexity is reflected on the complexity of the model, however; the model can be simplified by disintegrating it into small models. The smaller models are analyzed and validated and then integrated into the overall model analysis for the overall optimum performance. This principle is known as "Keep It Simple" (KIS). Building on this principle, in this chapter a system dynamic model involving human resource management for a multi unit maintenance system is discussed. Figure 3 shows, the CLD of maintenance system and partially the CLD of human resource management system.

It is argued that system dynamics modelling approach provides simplification to the complexity associated with developing a resources provision policy for multi unit maintenance system. The system dynamics model for multi unit maintenance system has been developed based on the CLD shown in Figs. 2 and 3 and then transformed into a system dynamics model as presented in Fig. 4. This model is reviewed and then verified using data from a selected case study.

4 Case Study

Case studies are suitable methods that can be used to verify the developed model for the maintenance resource policy. Afefy [15] in his chapter discussed the methodology, application of Reliability-Centered Maintenance (RCM) in a case

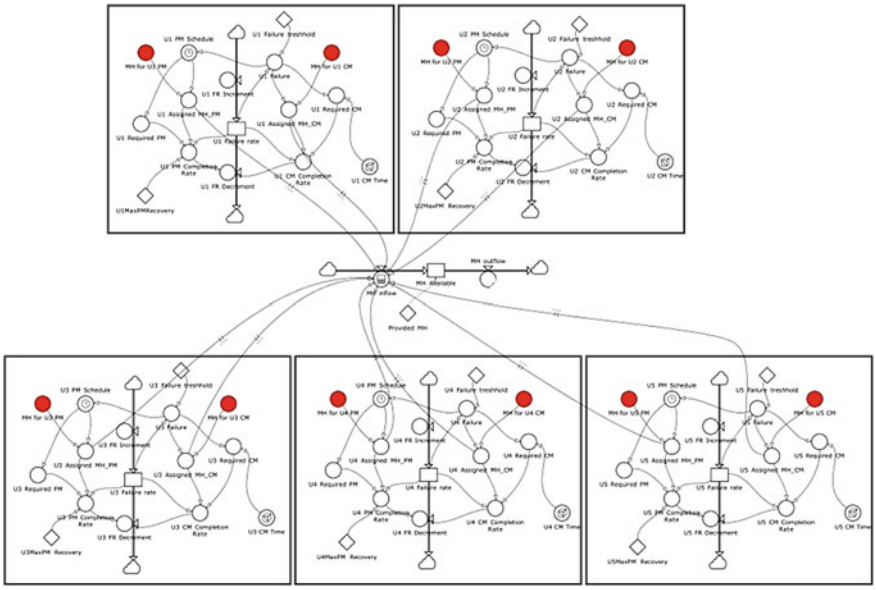


Fig. 4 System dynamics model for multi unit maintenance system

study. The result of the case study shows that the implementation of RCM can reduce labour cost significantly. This chapter adopts the same case study data to verify the implementation of the system dynamics model.

In this case study, there is a process-steam plant that consists of five different units of assets which are: fire-tube boiler, steam distribution, dryer, feed-water pump and process heater. These assets are respectively considered in this chapter as units 1, 2, 3, 4, and 5. The system dynamics model has been developed based on CLD presented in Fig. 2 but integrated all the five units of the case study. The complete system dynamics model used for this case study is presented in Fig. 4. Based on the case study, each unit has different down time and failure rate. Three different types of preventive maintenance system are implemented in each unit: weekly, monthly and six monthly. Each type of preventive maintenance requires different number of workers and duration which make up the man hours for the preventive maintenance activities.

One advantage of using a system dynamic simulation model is its ability to capture uncertain event from the real system. To capture the uncertain events, the input of simulation must be in a stochastic variable that can be represented in a certain distribution function. In the original case study, the number of required man hour is presented in a fixed number that is not convenient to capture the uncertainty. If the distribution is unknown, the best way is to assume it as a uniform distribution. So, for simulation purpose, required man hour is converted into uniform distribution. The result of the conversion is provided in Table 1 along with the corrective maintenance at the time of undertaking it.

Table 1 Man hours required for each maintenance activity

	Required MH for preventive maintenance			Corrective maintenance
	Weekly	Monthly	Six monthly	
Unit 1	Uniform (5, 7)	Uniform (8, 12)	Uniform (80, 88)	Uniform (29, 31)
Unit 2				Uniform (29, 31)
Unit 3				Uniform (9, 11)
Unit 4				Uniform (12, 21)
Unit 5				Uniform (9, 11)

In this case study, the total labour per day is 27 man day and it is assumed that a person work for 7 h per day. Therefore the total man hour available each day is 189 man hours per day. Another assumption used in this model is all people have the same ability as a maintainer.

Simulation is conducted in two different scenarios. First scenario is the current condition where there are 189 man hours provided. The second scenario is based on reducing the number of man hours relative to the result of the first scenario. The failure rate of each unit associated with these two different scenarios is compared.

The simulation is performed for 365 workdays to understand the behaviour of the multi unit maintenance system in the whole year. The result of the first scenario is shown in Figs. 5 and 6. Figure 5 shows the comparison between daily required man hours and available man hours. Figure 6 shows the daily failure rate of each unit.

Figure 5 shows that in average, there are significant numbers of man hour that are available. It designates that there is excess in the number of labour provided for the maintenance system. The simulation result also shows that the maximum required number of man hours in the whole simulation process happens in day 215 as much as 91.99 man hours. In Fig. 6, most of the daily failure rate is under 30 % but unit-1's failure rate is considerably different. Unit 1 (fire-tube boiler) experienced failure twice in the simulation time horizon. According to this result, the second scenario is developed with reducing man hours to 91 man hours and run the simulation for 365 days. The result is presented in Figs. 7 and 8.

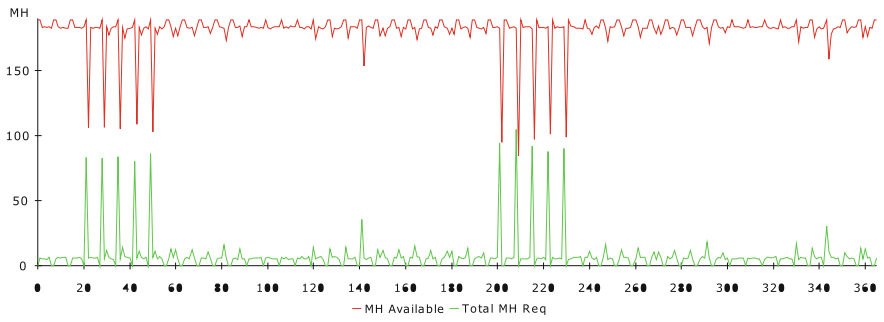


Fig. 5 Daily required and available man hours (scenario 1)

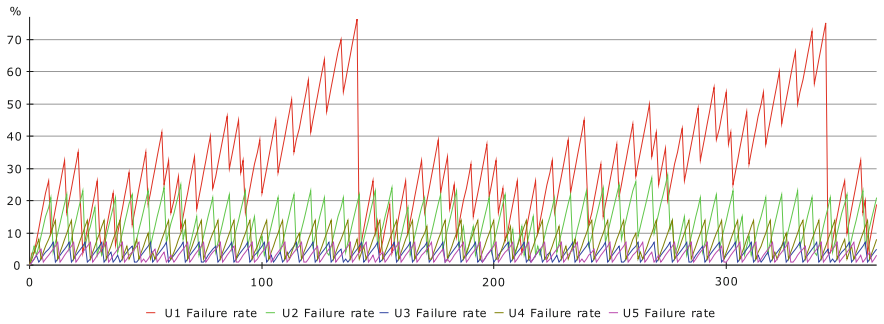


Fig. 6 Daily failure rate for all unit (scenario 1)

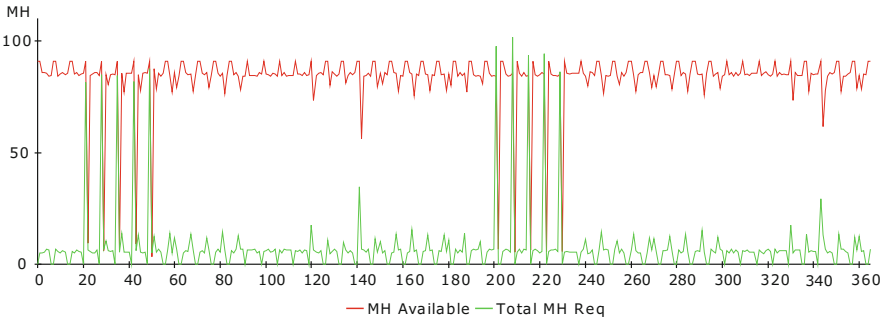


Fig. 7 Daily required and available man hours (scenario 2)

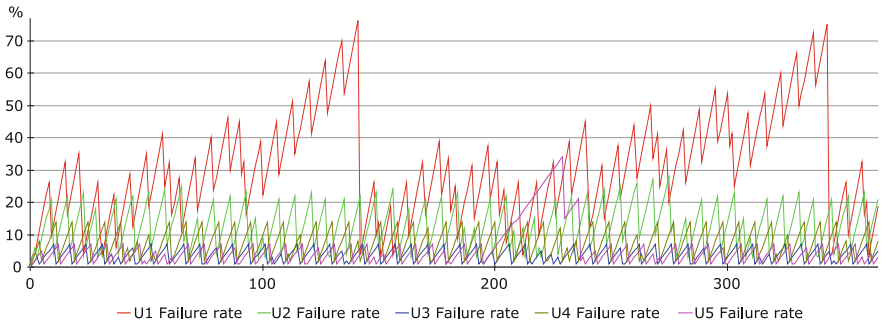


Fig. 8 Daily failure rate for all unit (scenario 2)

5 Discussion

Figures 7 and 8 show the result of the second scenario of the simulation model. It can be seen in Fig. 7 that although after 200 days the required man hour is more than the available man hour but overall the performance of the multi unit technical system is similar based on the daily failure rate. The significant difference can be found in unit 5 where the failure rate rises significantly after 200 days. This happens because at that period, the available man hour is insufficient to do preventive maintenance for all units. However, the man hour inadequacy to cover the preventive maintenance is only temporary and then the required preventive maintenance of unit 5 can be covered by available man hours in the next day. In this situation the shortage in man hours was resolved by delaying allocation of man hour for 1 day without causing failure or breakdown. So overall, considering the number failure and allocation of man hours led to proper management of the whole maintenance system.

6 Conclusion

Maintenance resources provision has a major role in asset management. More efficient maintenance resource provision process can lead to a better overall asset management performance. In this case, system dynamics model can be used to optimize the overall performance of multi unit maintenance system by determining the appropriate number of available resources to cover maintenance activities. Based on a case study data, a policy of maintenance resource provision is developed based on analyzing and comparing the output of current and suggested system dynamics model.

For the purpose of this chapter, the system dynamics model in this case study has only dealt with human resource management system for multi unit maintenance system, however; the mode can be implemented in include a wide range of resources. This chapter is one step in a research that aims at developing a more complex model that involves procurement of resources as well human resources management. This may require improvement of the model in terms of better man hours scheduling to smooth the requirement of daily man hours. Some adjustment might also be added to the model to accommodate outsourcing.

References

1. Ilyas Mohammed I, Cassady CR, Edward AP (2006) Establishing maintenance resource levels using selective maintenance. Eng Econ 51(2):99–114
2. Wang Y (2011) Predication and optimization of maintenance resources for weapon system. Int J Intell Syst Appl 3(5):1–9

3. Tsai YT, Wang KS, Tsai LC (2003) A study of availability-centered preventive maintenance for multi-componen system. Reliab Eng Syst Saf 84(3):261–270
4. Cui L, Li H (2006) Opportunistic maintenance for multi-component shock model. Math Method Oper Res 63(3):493–511
5. Okogbaa OG, Otieno W, Peng X, Jain S (Writer) (2008) Transient analysis of maintenance intervention of continuous multi-unit systems. IIE Transactions (Institute of Industrial Engineers, Inc. (IIE)) 40(10):971–983
6. Altiok T, Melamed B (2007) Simulation modeling and analysis with arena. Academic Press, San Diego
7. Endrenyi J, Aboresheid S, Allan RN, Anders GJ, Asgarpoor S, Billinton R et al (2001) The present status of maintenance strategies and the impact of maintenance on reliability. In: Paper presented at the IEEE transactions on power systems, New York
8. Tam ASB, Chan WM, Price JWH (Writer) (2006) Optimal maintenance intervals for a multi-component system. Prod Plann Control 17(8):769–779
9. Dwight R, Gordon P, Scarf PA (2012) Dynamic maintenance requirements analysis in asset management. Paper presented at the european safety and reliability conference, Troyes, France, 18–22 September 2011. CRC Press, Taylor & Francis Group, 847–852
10. Xiaohu Y, Xisen W, Yanling Q, Yongmin Y (2007) Modeling and simulation of complex maintenance system dynamics. In: Paper presented at the 26th Chinese control conference
11. Yang M, Zhu Y, Song J, Gu X (2009) A system dynamics approach for integrated decision making optimization of maintenance support system. In: Paper presented at the control and decision conference
12. Kothari V (2004) Assessment of dynamic maintenance management. Virginia Polytechnic Istitute and State University, Blacksburg
13. Bivona E, Montemaggiore GB (Writer) (2010) Understanding short- and long-term implications of "myopic" fleet maintenance policies: a system dynamics application to a city bus company. System dynamics review
14. Maani KE, Cavana RY (2007) Systems thinking, system dynamics: managing change and complexity. Pearson Education New Zealand, North Shore
15. Afefy IH (2010) Reliability-centered maintenance methodology and application: a case study. Engineering 2(11):863–863

Making Optimal and Justifiable Asset Maintenance Decisions

Andrei Furda, Michael E. Cholette, Lin Ma, Colin Fidge, Wayne Hill and Warwick Robinson

Abstract Maintenance decisions for large-scale asset systems are often beyond an asset manager's capacity to handle. The presence of a number of possibly conflicting decision criteria, the large number of possible maintenance policies, and the reality of budget constraints often produce complex problems, where the underlying trade-offs are not apparent to the asset manager. This chapter presents the decision support tool Justification and Optimisation of Budgets (JOB), which has been designed to help asset managers of large systems assess, select, interpret and optimise the effects of their maintenance policies in the presence of limited budgets. This decision support capability is realized through an efficient, scalable backtracking-based algorithm for the optimisation of maintenance policies, while enabling the user to view a number of solutions near this optimum and explore trade-offs with other decision criteria. To assist the asset manager in selecting between various policies, JOB also provides the capability of Multiple Criteria Decision Making. In this chapter, the JOB tool is presented and a real-world case study on a power plant system demonstrates JOB's capability to significantly improve the decision making process.

A. Furda (✉) · M.E. Cholette · L. Ma · C. Fidge
Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia
e-mail: andrei.furda@qut.edu.au

M.E. Cholette
e-mail: michael.cholette@qut.edu.au

L. Ma
e-mail: l.ma@qut.edu.au

C. Fidge
e-mail: c.fidge@qut.edu.au

W. Hill · W. Robinson
Delta Electricity, Wallerawang, NSW 2845, Australia
e-mail: wayne.hill@de.com.au

W. Robinson
e-mail: warwick.robinson@de.com.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_23

Keywords Asset management · Decision support · Combinatorial optimisation · Backtracking · Multiple criteria decision making

1 Introduction

Maintenance decisions for large-scale asset systems are often beyond an asset manager's capacity to handle. The presence of a number of possibly conflicting decision criteria, the large number of possible maintenance policies, and the reality of budget constraints often produce complex problems, where the underlying trade-offs are not apparent to the asset manager. For example, while one of the major objectives is to reduce costs by avoiding unnecessary maintenance work, another, conflicting objective is to improve the system availability and reliability.

Furthermore, in addition to clearly defined optimisation criteria, asset managers often rely on their experience, knowledge, and other external system-specific factors, which are very difficult to model in a purely optimisation-based tool. In contrast to optimisation tools, decision support tools provide not only optimisation results, but also allow asset managers to interactively select and analyse various alternatives, and also compare them to suggested optimal choices.

This chapter presents the decision support tool JOB (Justification and Optimisation of Budgets), which has been developed by CIEAM (Cooperative Research Centre for Infrastructure and Engineering Asset Management) and QUT (Queensland University of Technology). The JOB tool has been designed to help asset managers of large systems assess, select, interpret and optimise the effects of their maintenance policies in the presence of limited budgets. This decision support capability is realized through an efficient, scalable backtracking-based algorithm for the optimisation of maintenance policies, while enabling the user to view a number of solutions near this optimum and explore trade-offs with other decision criteria.

The remainder of this chapter is structured as follows. Section 2 gives an overview of related research, Sect. 3 presents the basic theory of the JOB tool, Sect. 4 elaborates on the application of JOB in a power plant case study. Section 5 concludes this chapter.

2 Related Work

System maintenance has evolved from a necessary productivity maintaining activity into an important business and asset management activity, resulting in significant research efforts dedicated to system maintenance. An overview study of the maintenance of complex systems can be found in [1] and [2], the topic of asset maintenance management is addressed in details in [3] and [4]. A maintenance decision support system overview is presented in [5], a specific solution suggestion using the

Analytical Hierarchy Process (AHP) and Fuzzy Logic is presented in [6], [7] suggests a multi-agent based approach. None of these publications have presented an effective system framework that is tested in a real business environment.

The approaches applied in the JOB decision support tool are well-founded and often applied in combinatorial optimisation and decision making. However, to the best of the authors' knowledge, their application for maintenance decision support, especially for Power Plant systems, is absent from published literature. Furthermore, one of the main aspects which make the JOB tool stand out is its advanced state of applicability in industry as a commercialization ready outcome of the research and utilisation. The JOB tool has been developed in cooperation with one of the largest electricity generators in Australia, specifically for maintenance decisions of power plant systems. Nevertheless, due to the advanced software architecture, its application is not limited to such systems, but can be extended to other complex systems of asset intensive industry.

3 Theory and Solution Methods of JOB

The presence of a number of possibly conflicting decision criteria, the large number of possible maintenance policies, and the reality of budget constraints often produce complex problems, where the underlying trade-offs are not apparent to the asset manager. JOB's purpose is to assist decision maker to select a maintenance option for each component of the complex system under a constrained budget, while optimising globally a number of (possibly competing) decision criteria. For example, the goal can be to reduce failure rates and the decision risk, while at the same time minimizing the system downtime due to maintenance and remaining within a specified budget.

The decision support functionality provided by the JOB tool relies on the solution of a number of (combinatorial) optimisation problems, where the maintenance options are the decision variables and the objective is to minimize various decision criteria that have business relevance, e.g. return on investment, failures per year, etc.

Each system component $i, i = 1, 2, \ldots, n$ has maintenance options $o_j^i, j = 1, 2, \ldots, m_i$ which indicate the type of maintenance, costs, needed time, etc. For example, a component could be replaced entirely, or only partially replaced, and repaired partially.

A maintenance policy, A, is defined by a selection of options for each system component $A = (a_1, a_2, \ldots, a_i, \ldots, a_n)$, where:

$$a_i \in \{o_1^i, o_2^i, \ldots, o_{m_i}^i\} \triangleq D_i$$

The set of all possible solutions is $A = D_1 \times D_2 \times \ldots \times D_n$ and thus $\mathrm{A} \in A$ is the set of all possible policies in the absence of additional constraints.

JOB solves the following optimisation problem:

$$\begin{aligned} & \min_{A} J_{DC}(A) \\ \text{subject to:} \quad & A \in A \\ & B_\ell \leq cost(A) \leq B_u \\ & LB_{DC_{limit}} \leq v_{DC_{limit}}(\mathrm{A}) \leq UB_{DC_{limit}} \end{aligned} \quad \text{(OPa)}$$

where DC is a decision criterion to be optimised, such as failure rate, number of outages, or the negative of return on investment. DC_{limit} denotes a decision criterion that is not being optimised and $v_{DC_{limit}}(\mathrm{A})$ denotes its value for the specified maintenance policy. $LB_{DC_{limit}}$ and $LB_{DC_{limit}}$ are the lower and upper bounds respectively, while B_ℓ and B_u denote the upper and lower maintenance budget. In other words, optimisation problem (OPa) is the minimization of a single decision criterion, with a constrained budget and additional constraints limiting value of other different decision criterion important to the decision maker. For example, the user might be interested in minimizing the failure rate, subject to a limited budget.

The set of *feasible* maintenance policies is denoted as

$$\begin{aligned} C &\triangleq A \cap C_{constraint} \\ C_{constraint} &= \{\mathrm{A} \mid B_\ell \leq cost(A) \leq B_u \text{ and } LB_{DC_{limit}} \leq v_{DC_{limit}}(\mathrm{A}) \leq B_{DC_{limit}}\} \end{aligned}$$

which is the set of all maintenance policies that satisfy *all* constraints in (OPa). This can now be stated in the compact form

$$\begin{aligned} & \min_{A} J_{obj}(\mathrm{A}) \\ & \text{subject to :} \quad \mathrm{A} \in C \end{aligned} \quad \text{(OP)}$$

Since each maintenance option of each system component can be combined with each maintenance option of a different component, this leads to a combinatorial problem with exponential complexity. The following sections address solution approaches.

3.1 Solution via Exhaustive Enumeration

The simplest way to solve this optimisation problem is a complete enumeration of all maintenance policies by generating all possible combinations of maintenance options and calculating the decision criteria values for each solution. This is also desirable for decision support, since one can examine any policy and its effect on any criterion. The number of policies in the solution space A is

$$|A| = \prod_{n=1}^{N} opt_n$$

where N denotes the number of system components and opt_n denotes the number of maintenance options for component n. Due to this combinatorial complexity, the complete enumeration of all solutions is feasible only for a system with a low number of constituent components, each with only a few options. Larger solution spaces will exceed the memory limitations if complete enumeration is pursued. The following section presents a solution to this problem.

3.2 Solution via Backtracking Based Algorithm

To remedy the problems related to memory and runtime limitations for systems composed with medium-to-large numbers of components, JOB utilizes a search strategy called *backtracking* [8] which explores the feasibility of partially-specified solutions prior to completing them.

When a partially-specified maintenance policy, $\hat{A} = (a_1, a_2, \ldots, a_k, \times, \ldots, \times)$ does not satisfy the problem constraints, the algorithm backtracks by unassigning a_k, since it was this assignment that resulted in the constraint violation. This backtracking discards entire sets of infeasible maintenance policies *en masse*, saving computational effort. If previously unexplored selections of a_k exist, the algorithm assigns one of these and the process is repeated. The algorithm terminates when the backtracking operation results in a completely unassigned maintenance policy and there are no unexplored maintenance options for the first component.

To solve the optimisation problem (OP) for any single criterion using backtracking, only the best (complete) policy at any given time has to be stored. Clearly, if the current, partially-specified policy cannot be completed in a manner that results in a better policy than this current optimum, backtracking should be invoked. However, to support more nuanced, multi-criteria analysis, it is desirable to be able to keep the best N maintenance policies for a single criterion. Unlike exhaustive enumeration, there is no guarantee that this N-policy set will contain the optima for any decision criteria other that the single criterion under consideration. However, the set will allow the decision maker to evaluate the sensitivity of other decision criteria in the neighbourhood of a single-criterion optimum, maintaining the decision support paradigm of JOB.

To store the best N policies while speeding the optimisation, a new *forward check* is defined. The definition of the current *stack* of the best policies is given by:

$$S = \{A_1, A_2, \ldots, A_N\}$$

Next, define a function that returns a lower bound on the partially-specified solution, $\hat{A}$

$$obj\_\min\left(\hat{A}\right) = \sum_{i=1}^{N} c_i$$

$$c_i = \begin{cases} obj(a_i) & \text{if } a_i \text{ is assigned} \\ \min\left(obj\left(o_1^{(i)}\right), \ldots, obj\left(o_{m_i}^{(i)}\right)\right) & \text{if } a_i \text{ is not assigned} \end{cases}$$

If

$$obj\_\min\left(\hat{\mathrm{A}}\right) < \max_{\mathrm{j}} J_{DC}\left(A_j\right)$$

then $\hat{A}$ *may* enter the stack of best solutions when it is completed. A similar forward check is conducted with respect to the budget upper and lower limits. Thus, if this forward checking condition (and similar checks on the budget upper and lower limits) are satisfied, the children of $\hat{A}$ are explored.

3.2.1 Multiple Criteria Optimisation

Multiple Criteria Decision Making (MCDM) is a widely applied discipline used to support decision makers deal with multiple and often conflicting objectives. MCDM problems are subdivided into two categories:

- Multiple Attribute Decision Making (MADM) techniques address discrete problems (i.e. the set of decision alternatives is discrete) [9, 10], while
- Multiple Objective Decision Making (MODM) techniques address continuous decision making problems [11].

The JOB tool supports discrete decision alternatives, and currently implements two of the most widely used decision making techniques:

- Simple Additive Method, and
- Weighted Product Method [10].

The subset of feasible maintenance policies (i.e. decision alternatives) is calculated by applying constraints on the set of alternatives provided by either exhaustive search, or the backtracking algorithm. Maintenance policies with decision criteria values outside the constraint limits are excluded from the set of feasible alternatives.

Each feasible maintenance policy is evaluated with respect to each decision criterion as follows. In a first step, the maximum and minimum are calculated and stored for each decision criterion.

Decision criteria are either of cost or of benefit type. Cost-type criteria are "better" for lower values, while benefit-type criteria are "better" for higher values. The goal is to minimize cost type criteria, and maximize benefit type criteria. This is taken into account when calculating the value function for each alternative and each decision criterion. Cost type criteria are expenditure, Decision Risk, failures, outages, while the return on investment (ROI) is a benefit criterion.

For each feasible maintenance policy, a utility value between 0 and 1 is calculated using an increasing linear function for benefit criteria, and a linear decreasing function for cost criteria. Weight factor parameters allow the user to indicate the importance of each decision criterion.

The Simple Additive Weighting Method calculates the value of a decision alternative i as:

$$V(Alternative_i) = \sum_{j=1}^{n} w_j * v_i^j$$

where w_j are weight factors, and v_i^j represent utility functions for a decision alternative i with respect to the decision criterion j.

The Weighted Product Method calculates the value of a decision alternative i as:

$$V(Alternative_i) = \prod_{j=1}^{n} (v_i^j)^{w_j}$$

3.3 User Interface and Process

The JOB software tool requires the user to specify a maintenance project containing a list of system components. Data about system components is: failure rates before maintenance, costs in the case of failures, estimated repair times, and age-related linearly increasing rates of failure. Maintenance options require information regarding maintenance cost, production loss, likelihood of failure after the maintenance, and the maintenance time.

A second user interface tab allows the user to specify decision constraints, MCDM parameters (see Sect. 3.2.1), and parameters for the backtracking-based combinatorial optimisation algorithm. A third user interface tab allows the decision maker to utilise either exhaustive enumeration or backtracking to optimise the system maintenance policy and presents the results. Also in this tab, the what-if analysis and report generation functionalities are available for the user to interpret and present the optimisation results.

Figure 1 shows the JOB decision support process. After creating or loading an existing data input file (Job Definition file), the user starts the analysis process. JOB calculates and displays a range of suitable maintenance policies, which can be selected and evaluated individually in terms of the following decision criteria (Fig. 2):

- Decision Risk, Failure Risk
- Total Number of Failures, Failures per Year
- Return on Investment, Availability
- Total Number of Outages, Outages per Year

Note that these criteria can be re-defined to suit other business specific applications. Results of a suitable user accepted maintenance policy can be stored in a report.

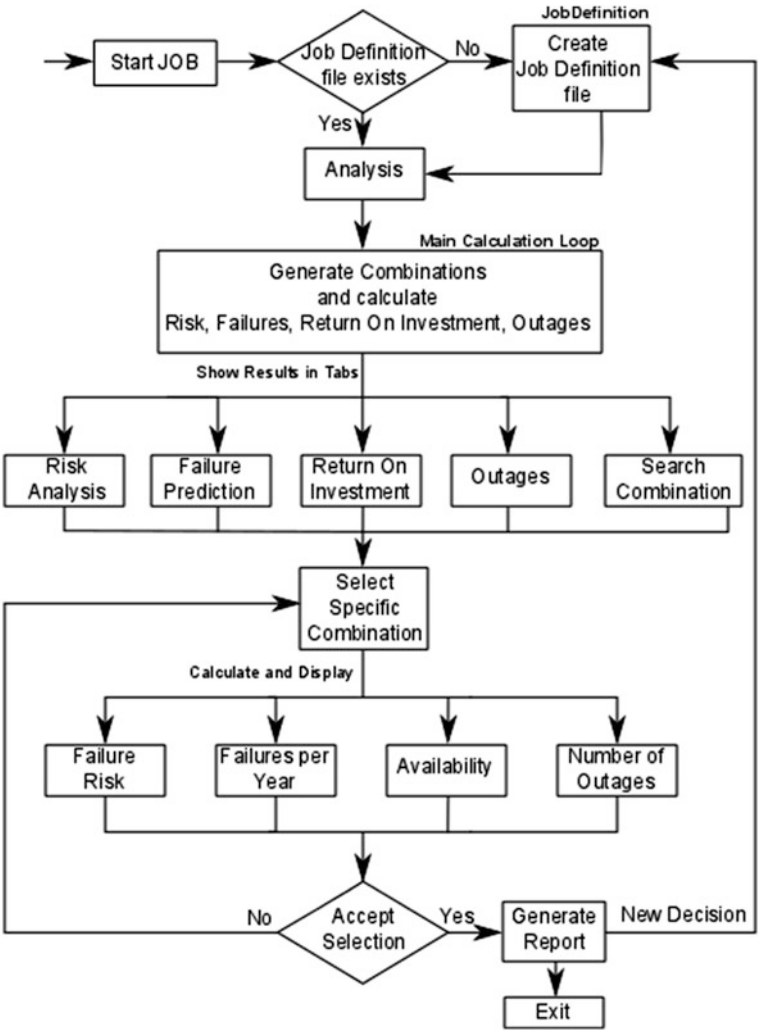


Fig. 1 JOB use case flowchart

4 Case Study: Maintenance Decision Support for a Power Plant

The JOB decision support tool has been developed in close cooperation with a large power generation enterprise in Australia, with the purpose of applying it in real decision scenarios for optimising the major maintenance of power plant systems.

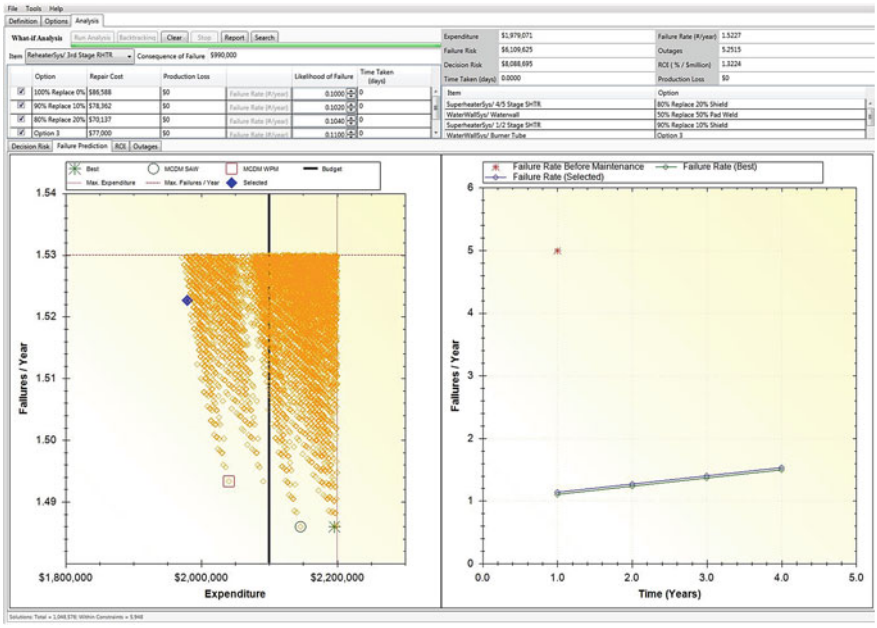


Fig. 2 JOB analysis and results user interface

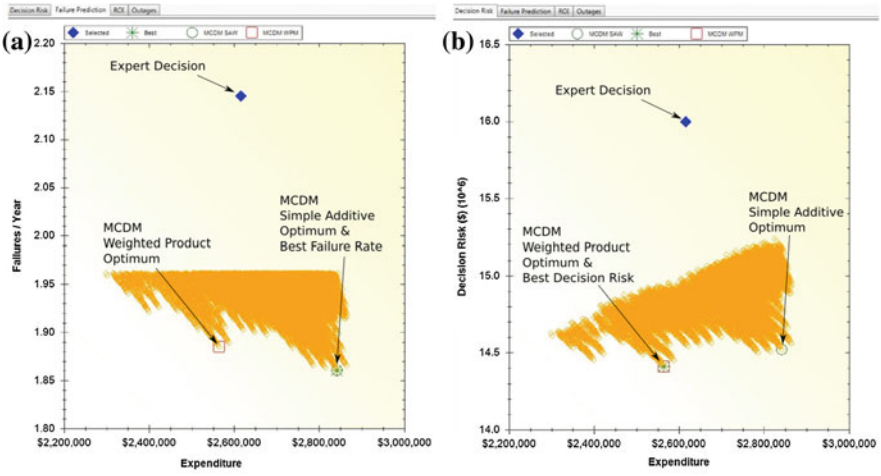


Fig. 3 Results of applying JOB to the minimisation of the number of failures per year. The best 20,000 maintenance policies (in terms of failures per year) are shown. The MCDM methods apply high weight factors for the failure rate criterion. **a** Failures per Year. **b** Decision Risk

A planned major power plant maintenance shut-down was used to evaluate the results provided by the JOB tool against the decision made by an expert human decision maker.

One of the test data sets for the evaluation was based on shut-down maintenance information of partial components of a large boiler system.

For this collection of systems, the number of possible maintenance policy combinations is approximately 33.5 million. Due to this large number, the back-tracking algorithm was utilised to generate maintenance policies while minimising the number of failures per year of the entire system. The results can be seen in Fig. 3a and b.

Depending on business and engineering constraints, there are a number of ways the decision could be improved. Some examples are:

- If additional dollars can be spent, the failure-optimal solution could be selected and a reduction of about 13 % failures per year could be achieved
- The expenditure could be kept *constant* and the failure rate could be reduced by about 0.5 failures per year by altering the selected maintenance options.
- The policy minimizing the decision risk could be selected. This policy would cost *less* money and incur less decision risk. Furthermore, this policy would have fewer failures per year.
- If the organisation decides that the current failure rate can be tolerated, the decision maker can decrease the maximum expenditures until the same number of failures per year is achieved by the optimised maintenance policy.

The MCDM Weighted Product Method parameterised with high weight factors for the failure rate (reflecting the asset manager's preference) results in a reduced expenditure of approximately 2 %, a reduced failure rate by approximately 13 %, and an increased return on investment in terms of availability return from 0.406 to 0.45 %/$million (Fig. 3).

It was found that by re-focusing the maintenance resources (selecting a different policy) the maintenance expenditures could be reduced to below $1.8 million while maintaining a failure rate below the baseline of 2.15 failures per year.

5 Conclusion

This chapter has presented JOB, an innovative maintenance decision support tool, which has been specifically developed to support asset managers with the challenge of large-scale maintenance decisions. Besides an exhaustive enumeration algorithm, the JOB tool solves the combinatorial optimisation problem using a highly scalable backtracking based algorithm, which allows the analysis of large systems in an efficient way. Furthermore, the JOB tool is able to suggest optimal maintenance policies based on Multiple Criteria Decision making techniques. A real-world case study attests its capability to support decision makers with complex maintenance decisions and its potential to save costs while even improving the system availability.

Acknowledgements This research was conducted within the CRC for Infrastructure and Engineering Asset Management, established and supported under the Australian Government's Cooperative Research Centres Program.

The authors would also like to thank the rest of the team members who contributed to the development of JOB: Dr. Yong Sun, Leon Mar, Andrew Sheppard, Lawrence Buckingham, and Dr. John Xie.

References

1. Murthy DNP, Kobbacy KAH (2008) Complex system maintenance handbook. Springer, London
2. Jardine AKS, Tsang AHC (2006) Maintenance, replacement and reliability: theory and applications. CRC Press, Boca Raton
3. Campbell JD (2010) Asset management excellence: optimizing equipment life-cycle decisions. CRC Press, New York
4. Hastings NAJ (2010) Physical asset management. Springer, London
5. Dong X-F, Gu Y-J, Yang K (2008) Study on intelligent maintenance decision support system using for power plant equipment. In: Automation and logistics, 2008. ICAL 2008. IEEE international conference on, 2008, pp 96–100
6. Tahir Z, Prabuwono AS, Burhanuddin MA, Akbar H (2008) Maintenance decision support fuzzy system in small and medium industries using decision making grid. In: Advanced computer theory and engineering, 2008. ICACTE '08 international conference on, 2008, pp 680–684
7. Yu-Jiong G, Xiao-Feng D, Jian-Jun W, Kun Y (2008) Study on multi-agent based maintenance decision support system used for power plant equipment. In industrial engineering and engineering management, 2008. IEEM 2008 IEEE international conference on, 2008, pp 2167–2171
8. Russell S, Norvig P (2003) Artificial intelligence: a modern approach. NJ: Pearson Education, Upper Saddle River
9. Tzeng G-H (2011) Multiple attribute decision making : methods and applications. Chapman & Hall/CRC, Hoboken
10. Yoon K, Hwang CL (1995) Multiple attribute decision making: an introduction, vol 07(104). CA Sage Publications, Thousand Oaks
11. Chankong V, Haimes YY (1983) Multiobjective decision making: theory and methodology, vol 8. North-Holland, New York

Assessment of Insulated Piping System Inspection Using Logistic Regression

Ainul Akmar Mokhtar, Nooratikah Saari and Mokhtar Che Ismail

Abstract Corrosion under insulation (CUI) is a common problem not only in chemical process plants but also in utility and power plants. According to empirical study, CUI is mainly driven by the operating temperature where CUI is more susceptible when the equipment or piping system is operating between −12 and 121 °C. Other factors such as insulation type and equipment or pipe location are also seen to be the contributing factors to CUI. However, to date, it is not clear which factors are more important in contributing to CUI occurrence. This paper presents a methodology for predicting the likelihood of CUI occurrence for insulated piping system using a logistic regression model. Logistic regression, a special case of linear regression, requires binary data and assumes a Bernoulli distribution. Using historical data, the variables of operating time in year, pipe operating temperature, type of insulation and pipe size are modelled as factors contributing to CUI. The outcome of this model does not produce the probability of failure to be used in quantitative risk-based inspection (RBI) analysis. However, the result rather uses the historical inspection data to provide the decision makers with a means of evaluating which pipe to be inspected for future planning of scheduled inspection, based on the likelihood of CUI occurrence.

Keywords Corrosion under insulation · Logistic regression · Likelihood of failure

A.A. Mokhtar (✉) · N. Saari · M.C. Ismail
Universiti Teknologi Petronas, Perak, Malaysia
e-mail: ainulakmar_mokhtar@petronas.com.my

N. Saari
e-mail: nooratikah_saari@petronas.com.my

M.C. Ismail
e-mail: mokhtis@petronas.com.my

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_24

1 Introduction

Corrosion under insulation (CUI) is one of the major problems for refineries, chemical and petrochemical process industries as well as for utility and power plants. In 1960s and 1970s, many plant designers were not concerned with the potential problems of CUI [1]. The consequence of the under design was many cases of pitting or rusting of carbon steel, stress corrosion cracking of austenitic stainless steel and other hidden metal loss found under the insulation. A study done by Exxon Mobil Chemical that was presented to the European Federation of Corrosion in September 2003 indicated that the highest incidence of leaks in the refining and chemical industries was due to CUI and not to process corrosion [2].

CUI is a severe problem because it can cause loss of production as well as affect the equipment or system integrity. Many chemical plants have experienced a variety of problems due to CUI [3]. Huge amounts of money have been spent for CUI inspection and maintenance where CUI contributed about 40–60 % of maintenance costs [2]. In one of the local companies, the maintenance cost for CUI has achieved approximately RM6 million which only covered CUI maintenance activities without including any other costs such as the non-destructive testing.

The current CUI inspection practice is that the qualified inspectors will inspect the insulated system visually for any susceptible sign for CUI based on several defined criteria such as the insulation and cladding quality (i.e. punctured, dislodged, missing or dented) or whether there is any sign of water ingress etc. Decision whether or not to open the insulation for further inspection will be based on the visual inspection result together with the factors outlined in the technical standards such as API 581. One of the factors is the operating temperature where equipment or piping system are more susceptible to CUI when operating between −12 and 121 °C [4]. Other factors such as type of insulation and equipment or pipe location are also the contributing factors to CUI. However, to date, it is not clear which factors are more important in contributing to CUI occurrence. For corrosion and inspection engineers, the difficulty that they are currently facing is to accurately make such decision in order to develop appropriate inspection and maintenance strategies using the available data. The visual inspection data recorded can be treated as binary data and can be further used to quantify the likelihood that the system will have CUI which can be employed for CUI inspection planning.

The objective of this paper is to present a methodology to assess the likelihood of CUI occurrence based on the visual inspection data. Logistic regression model is proposed in this study due to the nature of the visual inspection data that can be treated as binary data.

2 Corrosion Under Insulation

Corrosion is defined as a chemical or electrochemical reaction between material, usually a metal, and its environment that produces a deterioration of the material and its properties [5]. In other words, this means a loss of an electron of metal reacting with either water or oxygen. There are several types of corrosion and one of them is CUI. CUI is a localized corrosion occurring at the interface of a metal surface and the insulation on that surface. This can be a severe form of corrosion particularly because the corrosion occurs beneath the insulation. The process starts when there is water being trapped in between the metal and the insulation. The confined environment of the insulation material over the pipe, tank or equipment creates conditions that encourage build-up of moisture, resulting in corrosion. The corrosion is often more severe when the insulation restricts the evaporation process from occurring. In some cases the insulation acts as a carrier whereby moisture presents in one area moves through the insulation to another area causing the corrosion to spread more rapidly. Three factors are necessary for CUI to occur which are:

- *Water*: Water is the key point for corrosion to occur. Normally, water can be introduced from two sources, external and internal. Water infiltrates from external sources such as rainfall, steam discharge, spray fire sprinkles or drift from cooling tower. External water enters an insulation system through breaks or damages of the insulation which can happen during insulation storage and/or installation, through ineffective waterproofing, through maintenance or through service lapses. Even if the external sources are eliminated, water can still be introduced in the insulated system by the internal sources such as internal system leaks (e.g. water leak and steam tracing leak) or condensation. Condensation occurs when temperature of the metal surface is lower than the atmospheric dew point and causes poultice to trap in between metal and insulation.
- *Chemical content of water*: Chemical content of water plays an important factor for CUI to take place. Chlorides may be introduced by rainwater, plant and cooling tower atmospheres, misty sea environments or even portable water often used for fire-fighting, deluge testing or wash downs. Besides, traditional thermal insulation materials contain chlorides [2]. If they are exposed to moisture, chlorides released may form a moisture layer on the pipeline surface, resulting in corrosion.
- *Temperature*: Operating temperature also contributes to CUI. According to API 581, equipment or piping systems operating in the temperature range between −12 and 121 °C are more susceptible to CUI, with temperature range of 49–93 °C being the most severe environment.

3 Logistic Regression Model

Logistic regression models are extensively used in medical field [6–9]. For instance, Todd et al. [6] used logistic regression model to investigate the relationship between antioxidant vitamin intake and coronary heart disease in men and women. In social sciences, this model is broadly employed as well. Fuks and Salazar [10] applied logistic regression model to analyse the household electricity consumption classes. Paul [11] developed a logistic regression model to identify the various factors responsible for work related injuries in mines and to estimate the risk of work injury to mine workers. Other studies can be found in Refs. [12, 13].

Logistic regression also is widely used in business and marketing studies. For example, Sohn and Kim [14] provided a logistic regression model to predict the default of funded SMEs based on both financial and nonfinancial factors. Larivière and den Poel [15] studied the advantages for financial service providers in investing in youth marketing. Also, Cerpa et al. [16] developed a logistic regression analysis to predict the success rate of software development projects.

A review of the literature reveals the application of logistic regression model in analysing the dichotomous data to provide a basis for assessing systems subject to corrosion failure is limited. Spezzaferro [17] developed a logistic regression model to demonstrate the possibility of identifying statistical relationships between maintenance inspection interval lengths and corrosion observed percentages. The model provided a means for conducting trade-offs between inspection interval length and observed corrosion percentages in maintenance data, when measurable data are not available. Ariaratnam et al. [18] proposed a logistic regression model for predicting the likelihood that the sewer network, which is subject to corrosion failure as well, is in a deficient state.

Typically, for corrosion failure mode, the wall thickness data collected during inspection period are used to assess the probability of failure by analysing the data statistically. However, the wall thickness data is not always available for statistical methods to be used. Typically what is usually available in CUI inspection reports is the result from inspection after insulation removal which is corrosion was found and treated, or corrosion was not seen. These types of data are classified as binary responses with 0 and 1. Binary responses can be used to predict the probability of the occurrence using logistic regression model [19]. In statistics, logistic regression is used for prediction of the probability of occurrence of an event by fitting data to a logistic curve. It is a generalized linear model used for binomial regression. Like many forms of regression analysis, it makes use of several explanatory variables that may be either numerical or categorical.

It is important to understand that the goal of using logistic regression for data analysis is the same as that of any model-building technique used in statistics, that is, to find the best fitting and most parsimonious model. As in regression, a logistic regression model describes a relationship between a response and a set of explanatory variables. A response is also known as a dependent variable or an

outcome. Explanatory variables are also often referred to as covariates, independent variables or predictors.

To better explain the concept of logistic regression, the logistic function that describes the mathematics behind this regression should be defined. The logistic function, which ranges between 0 and 1, $f(z)$ is shown in Eq. (1). Plots of $f(z)$ yield an S-shaped curve resembling the cumulative distribution plot for a random variable.

$$f(z) = \frac{1}{1 + e^{-z}} \tag{1}$$

From the logistic function, the logistic regression model is obtained through the parameter z that can be written as the linear sum of the explanatory variables as Eq. (2).

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{2}$$

where $x_1, x_2, \ldots, x_n$ are defined as the independent variables of interest and $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ are the coefficient representing unknown parameters. Estimates of the parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ are obtained using a mathematical technique called maximum likelihood.

A regression can simultaneously handle both quantitative and qualitative explanatory variables. In the logistic regression model, the response variable is a binary variable whereas the explanatory variables can be either quantitative or qualitative variables. The quantitative variable can be further classified as a continuous variable, one that takes any value within the limits of variable ranges, for instance, the operating time. The quantitative variable can also be considered as a categorical variable such as operating temperature. For example, the pipe having operating temperature 290 °C can be categorized in group of pipes having operating temperature more than 121 °C.

The qualitative variable such as types of insulation is also considered as categorical covariates. Here, dummy variable needs to be used in order to overcome the weakness of the categorical variable as it cannot be meaningfully interpreted in regression model. Dummy variables are artificial explanatory variables in a regression model whereby the dummy codes are a series of numbers assigned to indicate group. In dummy variable, it will be binary variable as each variable is assumed one of two values, 0 or 1, indicating whether an observation falls in a particular group.

For dummy variable to be used, if there are K groups, one needs to have $K - 1$ dummy variables to represent K groups. Let say, if there are six operating temperature groups, one needs to have five dummy variables to represent the group which one of the groups will not be represented as dummy variable. It will be considered as a reference to which each of the group should be compared.

4 Methodology

In developing a logistic regression model, several steps have been employed as shown in Fig. 1.

Step 1: *Data Acquisition*
All information related to CUI was collected.
Step 2: *Variable definitions*
Define the response and explanatory variables.
Step 3: *Model development using MATLAB*

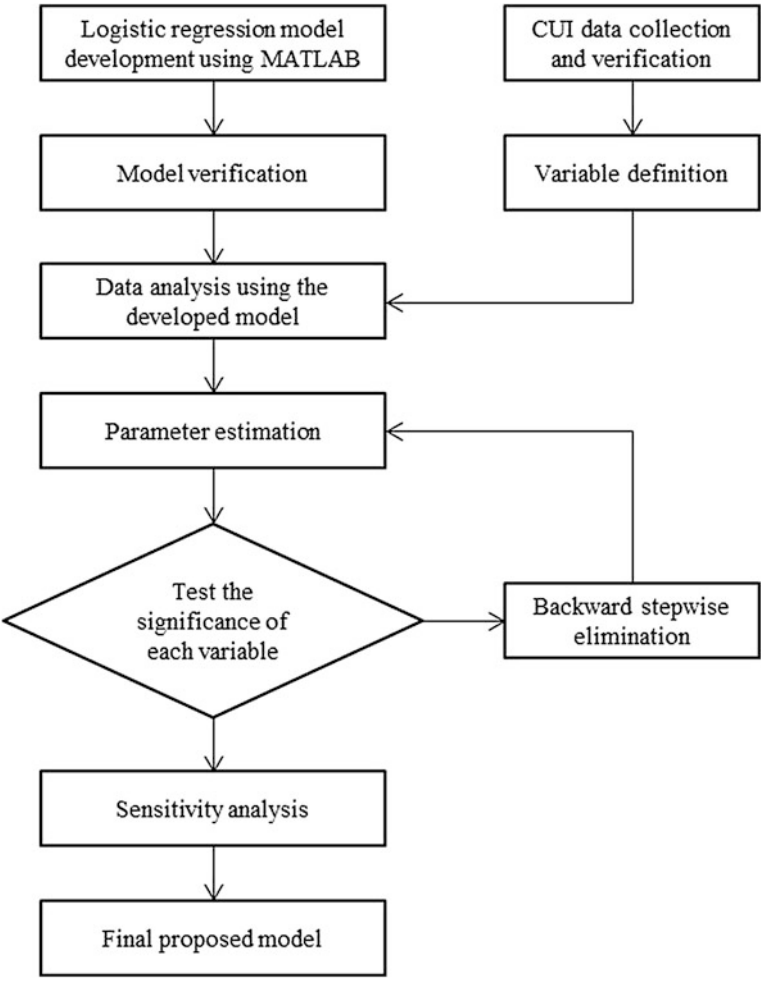


Fig. 1 Logistic regression model flowchart for CUI

The logistic model was developed in MATLAB R2009a using two main functions; *glmfit* and *glmval* functions. The *glmfit* function is used to estimate the coefficients of the parameters in the logistic regression model. The *glmval* function is used to compute the predicted values for the generalized linear model with link function '*link*' and predictors *X*.

Step 4: *Model verification*

To verify the logistic regression model developed using MATLAB, data taken from Evans Country case study was used [20]. Using the case study, the data was fitted to the logistic regression model using Java Script developed by Sullivan and Pezzullo [21]. The results generated by MATLAB were compared to the results generated via Java Script. The results showed that the logistic regression model developed in MATLAB produced the same coefficients as the results using Java Script by Kevin Sullivan. Thus, it is proved that the logistic regression model developed using MATLAB is acceptable.

Step 5: *Parameter Estimation*

Once a logistic regression is specified with its parameter and data have been collected, one is in a position to evaluate its goodness of fit, that is, how well it fits the observed data. Goodness of fit is assessed by finding the parameter value of a model and the procedure is known as parameter estimation [22].

Step 6: *Testing the significance of each parameter*

After generating the parameters for each variable, it is necessary to test the statistical significance of each parameter in the model. In the formulation of a logistic model, Wald test was performed on each variable or model parameter to investigate its significance [23]. Wald tests are based on the chi-square statistics that tests the null hypothesis that a particular variable has no significant effect given that the other variables are included in the model.

Step 7: *Backward Stepwise Elimination*

If the parameters obtained from result analysis are not significant, then a backward stepwise elimination method will be conducted to eliminate the parameter which is insignificant. Backward stepwise elimination method is an iterative variable-selection procedure where it begins with a model containing all the independent variables of interest. Then, at each step the variable with biggest *p*-value is deleted (if the *p*-value is bigger than the chosen cut-off level).

Step 8: *Sensitivity analysis*

The objective of the sensitivity analysis is to validate the proposed model and to test the reliability of the model by evaluating its sensitivity to minor changes in the data set. In order to conduct sensitivity analysis, new logistic models are developed using 80 and 90 % of the set of data based on the proposed method by Ariaratnam et al. [18]. Then, these new models will be compared to the 100 % data set (i.e. original data).

To show that there is no difference among these three models from statistical point of view, Kruskal-Wallis test was performed.

Step 9: *Estimate the probability of CUI occurrence to be used in RBI analysis* Once an appropriate model was validated, then the probability of CUI occurrence was estimated for inspection planning purpose.

5 Results and Discussions

CUI inspection data for small bore pipes in a gas processing plant were used to illustrate this methodology. Based on the data collected, explanatory variables were identified. In this study, three explanatory variables, which were pipe age, operating temperature and insulation type, were employed for the logistic model development based on the availability of data in the inspection database. Pipe age was classified as a continuous variable where the data came from three age groups (i.e. pipe age 6, 10 and 15 years).

Operating temperature was clustered based on API 581 operating temperatures and as such were defined as categorical variables. Categorical variables are the same as dummy variables which are artificial explanatory variables in a regression model. In this case, the dummy variables represent the categories of the operating temperature. Each dummy variable assumes one of the two values, 0 or 1, indicating whether an observation falls in a particular group. Operating temperature more than 121 °C was named as Group 6 and was referred as the reference group. Hence, five additional dummy variables were defined for operating temperature with respect to the reference as follow:

- Group 1: 1 when operating temperature is in the range 49–93 °C, 0 otherwise
- Group 2: 1 when operating temperature is in the range −12–16 °C, 0 otherwise
- Group 3: 1 when operating temperature is in the range 16–49 °C, 0 otherwise
- Group 4: 1 when operating temperature is in the range 93–121 °C, 0 otherwise
- Group 5: 1 when operating temperature is less than −12 °C, 0 otherwise

Insulation type can be classified into two groups and therefore, was also defined as categorical variables in the logistic model. There are two categories of insulation material: calcium silicate (type 1) and cellular glass (type 2). The response variable is classified as binary response and may be classified as either CUI is found ($Y = 1$) or CUI is not found ($Y = 0$).

The initial coefficients generated for small bore piping systems from MATLAB are shown in Table 1. It is observed that each of the coefficients are significant based on the p-value as it is lower than $\alpha = 0.05$ except for insulation type 1 (i.e. calcium silicate) which gave higher p-value. Thus, in this analysis, operating temperature showed a significant effect but insulation type gave no significance effect and may be removed from the model. Re-ran the analysis by excluding data

Table 1 Coefficients generated for small bore piping systems

	Initial run				Final (after re-run the analysis)			
Parameter	Coefficient	Standard error	Wald test	*p*-value	Coefficient	Standard error	Wald Test	*p*-value
Intercept	−4.1067	0.7491	−5.482	0.0000	−3.9804	0.6461	−6.161	0.0000
Pipe age	0.2365	0.0410	5.7683	0.0000	0.2366	0.0410	5.7701	0.0000
Temperature group								
Op. Temp. G1	1.9212	0.4529	4.2420	0.0000	1.8954	0.4458	4.2515	0.0000
Op. Temp. G2	1.7926	0.5634	3.1816	0.0015	1.6749	0.4404	3.8030	0.0001
Op. Temp. G3	1.5733	0.5314	2.9607	0.0031	1.4695	0.4317	3.4038	0.0007
Op. Temp. G4	1.6528	0.4482	3.6876	0.0000	1.6457	0.4473	3.6793	0.0002
Op. Temp. G5	1.3735	0.5284	2.5994	0.0093	1.2761	0.4433	2.8785	0.0040
Type of insulation								
Calcium silicate	0.1274	0.3775	0.3375	0.7357	n/a			

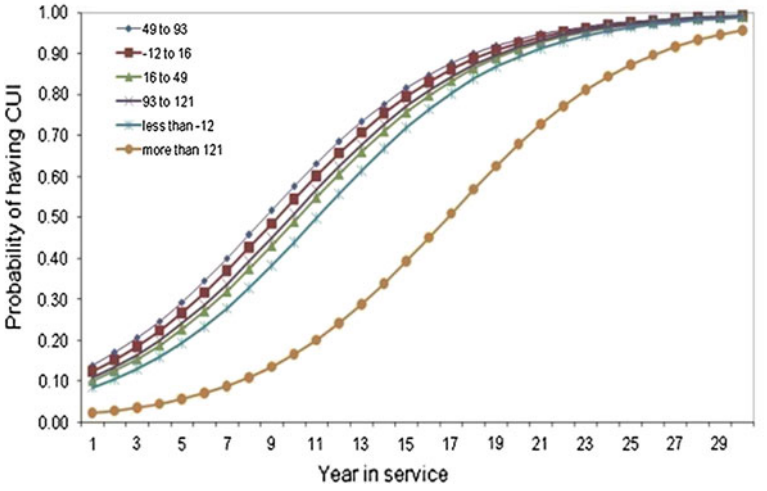


Fig. 2 Probability of CUI occurrence for each temperature group for small bore pipe

on insulation type using the backward stepwise elimination method yields the following results as shown in Table 1.

The logistic regression coefficient β_1 for pipe age is 0.236 with $\exp(\beta_1) = 1.267$. This implies that, when pipe age increases by 1 year, the likelihood of small bore pipe will have CUI increases by 26.68 %. The effect of ranges of operating temperature is as shown in Fig. 2. The probability for CUI occurrence for six groups of operating temperature for both small pipes was plotted against pipe age. The trend produced replicated the API guidelines where operating temperature group 1 (49–93 °C) showed the highest probability of having CUI when compared to other temperature groups.

From Table 2, all p-values have shown significant values as the values were lower than $\alpha = 0.05$. Thus, a general equation of a linear function of independent variables for small bore piping systems can be written as

$$y(x) = -3.980 + 0.237x_1 + 1.895x_2 + 1.675x_3 + 1.470x_4 + 1.646x_5 + 1.276x_6 \tag{3}$$

where x_1 = pipe age (years in service); $x_2, \ldots, x_6$ = dummy variable for operating temperature groups.

Let p_i be the probability of CUI occurrence in case i and the logistic regression model is [19]

$$\text{log it}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_1 \tag{4}$$

Table 2 Coefficients for 100, 90 and 80 % of sample data (small bore piping)

	100 % of sample data		90 % of sample data		80 % of sample data	
Variable	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
Intercept	−3.9804	0.0000	−4.0380	0.0000	−4.0624	0.0000
Age	0.2366	0.0000	0.2334	0.0000	0.2419	0.0000
Op. Temp. 1	1.8954	0.0000	2.1649	0.0000	1.9004	0.0002
Op. Temp. 2	1.6749	0.0001	1.7502	0.0001	1.7586	0.0002
Op. Temp. 3	1.4695	0.0007	1.6522	0.0003	1.3076	0.0046
Op. Temp. 4	1.6457	0.0002	1.7561	0.0001	1.8386	0.0004
Op. Temp. 5	1.2761	0.0040	1.3631	0.0036	1.0156	0.0405

The logistic regression model presents the log odds of CUI occurrence as a linear function of pipe age with respect to operating temperature group. To predict the probability of CUI occurrence at certain years in service, the proposed model is Eq. (5) by rearranging Eq. (4).

$$p = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}} \tag{5}$$

For example, the likelihood of CUI occurrence for the pipe operating between 49 and 93 °C (i.e. Group 1) and the age of pipe is 10 years old is

$$p = \frac{e^{-2.085+0.237(10)}}{1 + e^{-2.085+0.237(10)}} = 0.570$$

This means after 10 years of pipe being in service, there is a 57 % chance that the pipe will have CUI when the insulation is removed.

5.1 Sensitivity Analysis of Model

A sensitivity analysis was also performed to validate the proposed model. The logistic model was developed using two scenarios which are using 80 and 90 % of the data set based on the proposed method by Ariaratnam et al. [18]. The coefficients generated by the three groups of data are given in Table 2. The sensitivity analysis revealed that $KW = 0.089$ compared with $\chi^2_{0.05,2} = 5.991$. Therefore, the null hypothesis can be accepted, indicating that there is no significant difference among the three models. The proposed model seems to be a good representation of the observed data.

6 Conclusions

This study produced a mathematical model that provides the likelihood of having CUI for an insulated piping system given the pipe age, operating temperature and insulation type. These CUI factors have been discussed extensively in the literature but no mathematical model has been developed to show the relationship between the likelihood of having CUI and its factors. The results revealed that age and operating temperature have a significant effect on the deterioration of the small bore piping systems.

Intuitively, one knows that the likelihood of having CUI will increase as pipe aging. However, in API 581, the time factor is not being discussed explicitly. The logistic regression has managed to include time as one of the significant model parameters where the probability value can be obtained on certain year in service. Like operating temperature and insulation type, both are the factors for CUI discussed in API 581; nonetheless, the discussion is more towards a guideline. The logistic regression produced a mathematical model to that quantifies the likelihood of having CUI given both factors (i.e. operating temperature and insulation type).

References

1. Kletz TA (1995) Equipment maintenance. In: Grossel SS, Crowl DA (eds) Handbook of highly materials handling and management. Marcel Dekker Inc, New York, pp 439–498
2. Corrosion under insulation (2012) Corrosion under insulation (CUI): A nanotechnology solution. http://www.nansulate.com/pdf/CUI_nanotechnology_solution.pdf
3. Pollock WI, Barnhart JM (1985) Corrosion of metal under thermal insulation. American Society for Testing and Material, United States
4. Davis RJ (2000) Corrosion understanding the basic. ASM International, USA
5. American Petroleum Institute (2000) Base resource document on risk-based inspection. API Publication, Washington, DC, p 581
6. Todd S, Woodward M, Bolton-Smith C (1995) An investigation of the relationship between antioxidant vitamin intake and coronary heart disease in men and women using logistic regression analysis. J Clin Epidemiol 48(2):307–316
7. Ottenbacher KJ, Smith PM, Illig SB, Linn RT, Fiedler RC, Granger CV (2001) Comparison of logistic regression and neural networks to predict rehospitalization in patients with stroke. J Clin Epidemiol 54:1159–1165
8. Camdeviren HA, Yazici AC, Akkus Z, Bugdayci R, Sungur MA (2007) Comparison of logistic regression model and classification tree: an application to postpartum depression data. Expert Syst Appl 32:987–994
9. Austin PC, Tu JV, Lee DS (2010) Logistic regression had superior performance compared with regression trees for predicting in-hospital mortality in patients hospitalized with heart failure. J Clin Epidemiol. doi:10.1016/j.jclinepi.2009.12.004
10. Fuks M, Salazar E (2008) Applying models for ordinal logistic regression to the analysis of household electricity consumption classes in Rio de Janeiro. Brazil Energy Econ 30 (30):1672–1692
11. Paul P (2009) Predictors of work injury in underground mines: An application of a logistic regression model. Min Sci Technol 19:282–289

12. Can T, Nefeslioglu HA, Gokceoglu C, Sonmez H, Duman TY (2005) Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. Geomorphology 72:250–271
13. Sin SJ, Kim K (2008) Use and non-use of public libraries in the information age: a logistic regression analysis of household characteristics and library services variables. Libr Inf Sci Res 30:207–215
14. Sohn SY, Kim HS (2007) Random effects logistic regression model for default prediction of technology credit guarantee fund. Eur J Oper Res 183:472–478
15. Larivie`re B, den Poel DV (2007) Banking behaviour after the lifecycle event of "moving in together": an exploratory study of the role of marketing investments. Eur J Oper Res 183:345–369
16. Cerpa N, Bardeen M, Kitchenham B, Verner J (2010) Evaluating logistic regression models to estimate software project outcomes. Inf Softw Technol 52:934–944
17. Spezzaferro K (1996) Applying logistic regression to maintenance data to establish inspection intervals. In: Proceedings annual reliability and maintainability symposium, IEEE, pp 296–300
18. Ariaratnam ST, El-Assaly A, Yang Y (2001) Assessment of infrastructure inspection needs using logistic models. J Infrastruct Syst 7(4):160–165
19. Hosmer DW, Lemeshow S (1989) Applied logistic regression. Wiley, New York
20. Kleinbaum DG, Kupper LL, Morgenstern H (1982) Epidemiologic research: principles and quantitative methods. Van Nostrand Reinhold, New York
21. Sullivan, K. and Pezzullo, J.C. (2007). *Logistic Regression version 05.07.02*. Retrieved April 12, 2010, from http://statpages.org/logistic.html
22. Agresti A (1990) Categorical data analysis. Wiley, New York
23. Yang J, Gunaratne M, Lu JJ, Dietrich B (2005) Use of recurrent Markov chains for modeling the crack performance of flexible pavements. J Transp Eng 131:861–872

Continuous Life Cycle Cost Model for Repairable System

Masdi Muhammad, Meseret Nasir, Ainul Akmar Mokhtar and Hilmi Hussin

Abstract Traditionally, the estimation of maintenance cost of a repairable system was evaluated using discrete approach based on estimated number of system failure, cost of repair as well as the interest rates. As maintenance cost represents a significant portion of overall life cycle cost (LCC), accurate estimation of maintenance cost would influence LCC analysis. However, in actuality the failure of the repairable system occurs in a continuous probabilistic manner thus the assumption of discrete occurrence is rather inaccurate. This paper presents an alternative continuous LCC model to better represent the actual operating phenomena of repairable system. The model was established based on the widely used Weibull distribution probability density function and continuous combined interest method. The result of the developed LCC model was then validated using Monte Carlo method. The result indicates that the continuous LCC model is able to accurately estimate LCC for any given time that can be useful in decision making based on life cycle cost.

Keywords Life cycle cost · Repairable system

1 Introduction

Life Cycle Cost (LCC) has been an increasingly important criterion in decision making to evaluate alternatives of repairable systems, which could influence the performance of the business. In LCC analysis, all types of costs associated with a system's life cycle, starting from the acquisition, operation and maintenance and decommission costs are all taken into account [1]. LCC analysis is about keeping the operational level of asset with minimum cost [2]. Specifically for repairable systems, the operation and maintenance costs accounts for 70 % of total costs thus

M. Muhammad (✉) · M. Nasir · A.A. Mokhtar · H. Hussin
Universiti Teknologi Petronas, Perak, Malaysia
e-mail: masdimuhammad@petronas.com.my

M. Nasir
e-mail: meseretnasir@yahoo.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_25

warrant an accurate estimate. The maintenance of repairable system can be broadly categorized into two: preventive maintenance and breakdown maintenance. While the planned preventive maintenance's cost can be estimated relatively easily, the unplanned breakdown maintenance represents different level of challenge due to the probabilistic nature of failure occurrence. Unexpected failure of a repairable system can result in high costs for repair, replacement as well as the consequence of such failure [3]. The expenses due to unplanned breakdown are usually significantly contributed by the consequential cost of failure and thus will highly influence the outcome of LCC analysis. For instance, these costs can be over $300 billion on plant maintenance and operations in U.S. annually [4] and about 80 % of the expenses are spent to correct catastrophic failures of repairable systems [3].

Traditionally, the net present values (NPV) of maintenance and repair costs are estimated using discrete approach [5]. Zhu et al. on the other hand, presented a comparison between deterministic and probabilistic LCC of ground source heat pump [6]. Their result indicated that probabilistic method may not change the outcome of LCC but may change the sensitivity of the cost factor. This is due to the fact that the probabilistic part was introduced in the total cost estimate and not at the cost elements. The estimate of maintenance and repair costs, which is one of the critical cost elements, is based on the number of repair occurrences of the system over the period of interest [7]. However, stochastic nature of failure implies that the failure can occur randomly at any time during the life span of the repairable system. On this ground, the estimation of repair and maintenance cost using a discrete approach will not be able to accurately represent the actual condition. This paper focuses on the development of continuous time LCC model incorporating the stochastic nature of unplanned maintenance and repair cost. Due to the versatility of Weibull distribution in modelling the probabilistic nature of the failure, the distribution and the continuous combined interest method is adopted in estimating the NPV of the unplanned maintenance cost.

This paper is organized as follows. Section 2 presents the methodology used in deriving the governing equations with the inclusion of both failure probability and continuous compounding interest function. Section 3 discusses the results of the proposed method and compared against the discrete method. The section also elaborates the sensitivity of the model towards the changes in the Weibull's shape factor and the nominal interest rates. Lastly, the paper is concluded with a brief summary in Sect. 4.

2 Methodology

Generally, LCC in terms of net present value can be expressed as [8]:

$$\text{LCC} = \text{Acquisition Cost} + \text{Ownership Cost} + \text{Disposal Cost}$$

The ownership cost includes the cost of operation and maintenance and the estimates of these costs will be the focus of this research. In order to estimate the unplanned maintenance cost, there are two important parameters to be considered. These parameters are cost of repair including spare part cost, labour cost and consequential cost of the failure and the probability that the failure will occur. Based on these two parameters, the cumulative maintenance cost of the repairable system can be estimated using Eq. (1).

$$\Delta C_f = C_r f(t) K \Delta t \tag{1}$$

where

C_f is the cumulative cost of failure,
C_r is the cost of repair per failure,
$f(t)$ is the failure probability of the system and
k is the present value (PV) factor of cost.

2.1 Probability of Failure

The probability of failure is estimated with the assumption that the system is repaired to be as good as new following perfect renewal process. With these assumptions, the random variable of time between failures can be analysed using distribution fitting method in which Weibull distribution is used. The cumulative distribution function (CDF) of Weibull is given by;

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \tag{2}$$

where, $t \geq 0$, $\beta > 0$ is a shape parameter and $\alpha > 0$ is a scale parameter.

The probability density function (PDF) is given by;

$$f(t) = \frac{dF(t)}{dt} = \frac{\beta}{\alpha^{\beta}} t^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \tag{3}$$

Maximum likelihood function is then applied to estimate the parameters, β and α. The detail mathematical formulation of Maximum likelihood function can found in [9]. The likely hood function given by

$$L = \prod_{i=1}^{n} \frac{\beta}{\alpha^{\beta}} t^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \tag{4}$$

Taking the logarithms of (4) and differentiating with respect to β and α in turn and equating to zero, one can formulate Eqs. (5) and (6).

$$\frac{\partial L}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^{n} \ln t_i - \frac{1}{\alpha} \sum_{i=1}^{n} t_i^{\beta} \ln t_i = 0 \tag{5}$$

$$\frac{\partial L}{\partial \beta} = -\frac{n}{\alpha} + \frac{1}{\alpha^2} \sum_{i=1}^{n} t_i^{\beta} = 0 \tag{6}$$

where

t_i is the time between failures

n is the sample size

β and α can then be estimated by solving simultaneous Eqs. (5) and (6).

2.2 Cost of Repair

Maintenance or repair is the process of restoring the system back into operational state with associated cost including spare part cost, labor cost and consequential cost of the failure (loss of production, reputation). This cost can be incurred anytime in the life span of the repairable system with the occurrence of failure. The present value of the change in cost of repair can be estimated with Eq. (7) which assumes continuous compound interest rate.

$$\Delta C_r = C_r e^{-rt} \Delta t \tag{7}$$

where r is the nominal interest rate.

2.3 Total Maintenance Cost

By substituting Eqs. (3) and (7) into (1), the change in maintenance cost can be formulated as shown in Eq. (8)

$$\Delta C_f = C_r \frac{\beta}{\alpha^{\beta}} t^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta} - rt\right] \Delta t \tag{8}$$

By integrating Eq. (8), the cumulative present value of the maintenance cost any particular time can be estimated as in Eq. (9).

$$PVC_f = \int_0^t C_r \frac{\beta}{\alpha^\beta} t^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^\beta - rt\right] dt. \tag{9}$$

3 Results and Discussions

To illustrate the application of the continuous model, a numerical example is presented with constant cost of repair for the 20 years period. Other elements used in the example are as shown in Table 1. The result in terms of cumulative present value is shown in Fig. 1.

The cumulative cost of repair is increasing until it reaches the maximum which indicates the maximum contribution from the continuous discount rate factor and actual cost of repair for the particular year. After the 10th year, the contribution from the discount rate factor decreases rapidly overcoming the increasing probability of failure (or the number of failures). The relationship between the discount rate and the cumulative failures is shown in Fig. 2.

In order to validate the hypothesis, a Monte Carlo (MC) method similar to that proposed by Barringer [8], was employed as comparison. The MC algorithm can be surmised as follows:

1. Randomly generate the time between failures based on the given distribution. In this particular case, sample size is 20.
2. Calculate the cumulative time to failure (or the age at which the failure occurs).
3. Round-up the age to the next integer (i.e. failure at age 4.04 is round up to age of 5 years).
4. Repeat step 1–3 until the required number of iteration (1,000 iteration for this case).
5. Sum up the repair for each year for all the iteration.

Table 1 Parameters for numerical example

Parameters	Values
Repair cost	$10,000
Weibull distribution parameters	$\beta = 1$, $\alpha = 3$ years
Study period	n = 20 years
Nominal interest rate	10 %

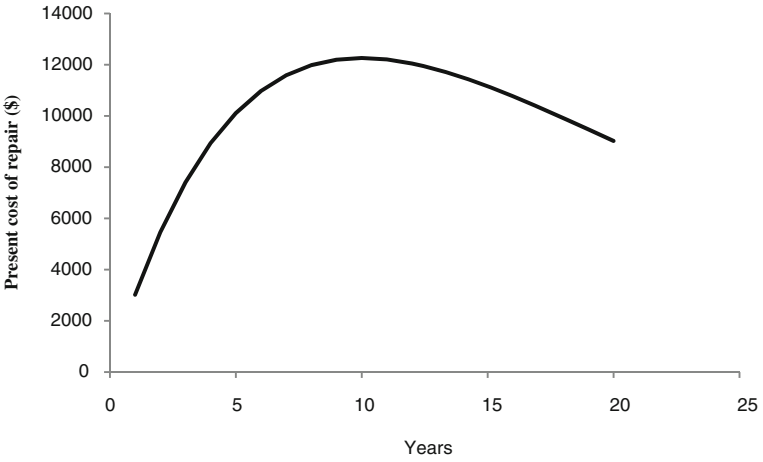


Fig. 1 Changes in present value of repair cost

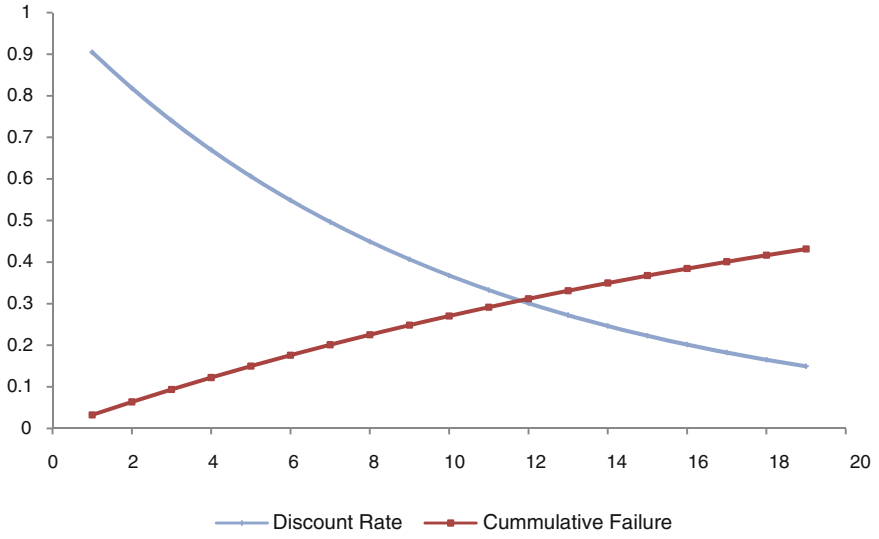


Fig. 2 Relationship between discount factor and cumulative failure

6. Calculate the repair cost by multiplying the number of repair with the cost of repair.
7. Calculate the average cost by dividing the total cost with the number of iteration.

In MC method, the present value of the repair cost is estimated by multiplying the number of failure for each year by the cost of repair and the discrete discount rate. The main objective of the comparison is to look at the difference in the

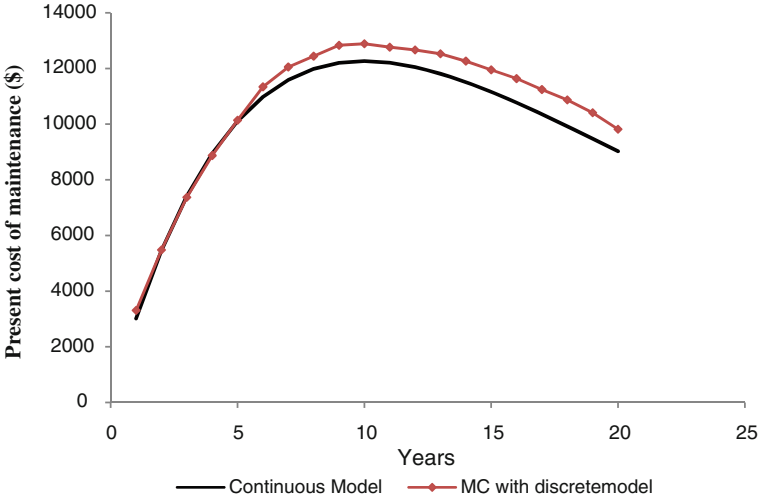


Fig. 3 Comparison of cumulative cost between continuous and discrete model

cumulative cost to the proposed continuous model. The result based on the numerical example is as shown in Fig. 3. The result of the study indicated that for the first 5 years both approaches estimated almost similar value of maintenance cost. However, as the period passed the eighth year, there is an increasing gap between the two models where at the end of 20 years, there is a maximum of 9 % gap in the cost estimates. The overestimation in the discrete model was mainly contributed by the higher expected number of failures, thus the higher cost. On the other hand, the continuous model represents the actual number of failure through analytical solution.

It is also interesting to note that with continuous model, the different region of the conceptual bathtub curve profile for repairable system can be assessed. The bathtub curve, which are categorized into three different stages according to the failure rates (decreasing, constant and increasing) can be defined based on the value of values of the shape parameter, β. For $\beta > 1$, indicating increasing failure rates, the maintenance cost will be low at the start of period but will be increasing with the increased length of operation. Meanwhile, with $\beta < 1$, indicating decreasing failure rate the maintenance cost will be high maintenance cost at the early stage of the operation and reducing with the increase period of operation. Figure 4 shows the change in the present value of maintenance cost with the changes in β values. Thus, this result proves that continuous model can be applied to estimate the maintenance cost at different state of the system life cycle.

Figure 5 shows the effect of the nominal interest rate on the present cost of maintenance. The result indicates that as the nominal interest increases, the present value of the maintenance cost reduces accordingly. The similar result can also be observed for discrete model. This implies the selection of interest rates do have the same impact to the continuous model as in the discrete model.

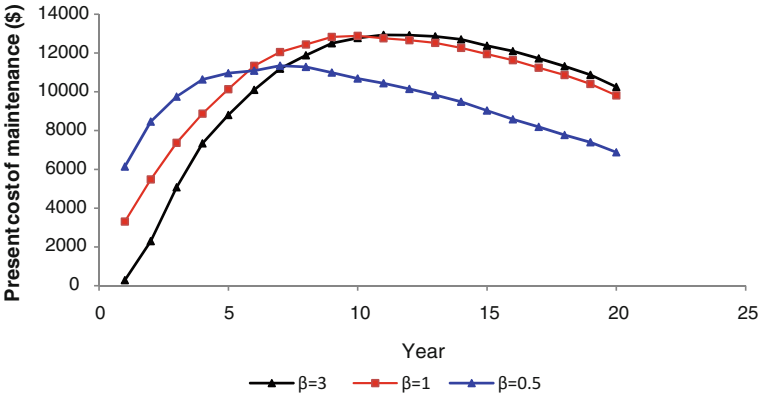


Fig. 4 Estimation of present cost of maintenance at different stages of repairable system

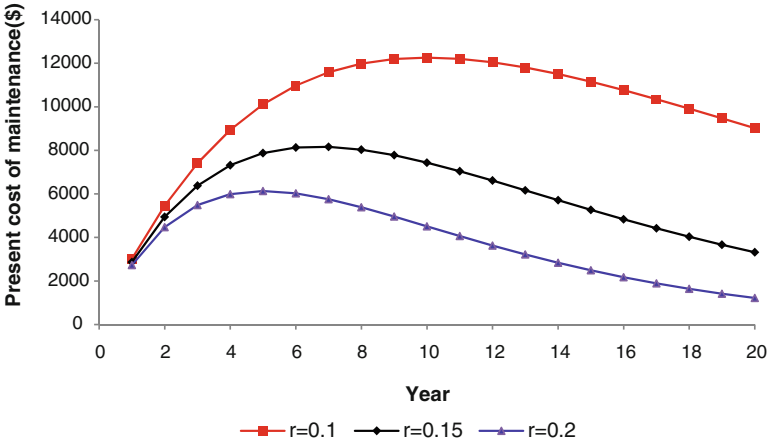


Fig. 5 Effect of nominal interest rate in the present cost of maintenance repairable system

4 Conclusion

This paper presented a continuous LCC model to estimate the LCC of repairable system specifically focusing on the probabilistic nature of unplanned maintenance cost. The findings of the study shows that the continuous model can be used to predict the maintenance cost contribution to LCC more accurately compared with discrete approach by taking into account the time to failure distribution and continuous discount rate factor. However, the results also indicate that the discrepancy between the two models is more significant with the longer period of study. For study period of less than 8 years, the results between the two methods are comparable.

The continuous model can also be expanded to include other time to failure distributions as well as to include the repair factor which could have high influence to the final cost of repair.

Acknowledgments The authors would like to acknowledge the support of Universiti Teknologi PETRONAS for this study.

References

1. Korpi E, Ala-Risku T (2008) Life cycle costing: a review of published case studies. Manag Audit J 23:240–261
2. Woodward D (1997) Life cycle costing–theory information acquisition and application. Int J Proj Manag 15:335–344
3. Khirshnasamy L, Khan F, Haddara M (2005) Development of risk-based maintenance (RBM) strategy for a power-generating plant. J Loss Prev Process Ind 18:69–81
4. Dhillon B (2002) Engineering maintenance: a modern approach. CRC Press, LLc, Boca Raton
5. Christiansen T (2013) Risk-based assessment of unplanned outage events and costs for combined-cycle plants. J Eng Gas Turbines Power 135:1–021801
6. Zhu Y, Tao Y, Rayegan R (2012) A comparison of deterministic and probabilistic life cycle cost analyses of ground source heat pump (GSHP) applications in hot and humid climate. Energy Build 55:312–321
7. Hennecke FW (1999) Life cycle costs of pumps in chemical industry. Chem Eng Process 38:511–516
8. Barringer PH (1996) Life cycle cost tutorial. In: fifth international conference on process plant reliability. Houston, Texas
9. Guure CB, Ibrahim NA (2013) Methods for estimating the 2-parameter Weibull distribution with Type-I censored data. Res J Appl Sci Eng Technol 5:689–694

Research on Sequencing Optimization of Military Aircraft Turnaround Activities Based on Genetic Algorithm

Boping Xiao, Shuli Ma, Haiping Huang and Aoqing Wang

Abstract Turnaround is a series of effective activities, which are prepared in order to turn out immediately next time after the last turn out according to the scheduled preparation project, turnaround time is one of the most important major parameter which can measure and judge the war power of the military aircraft [1]. In order to shorten the turnaround time, enhance the availability of the logistics resources, and reduce the vacancy rate of the support equipment and support personnel, in this paper, it applies activity route optimization algorithm of Genetic Algorithm (GA) in this article to analyze and research that turnaround activities of one particular type of task of military aircraft [2]. At first, these certain turnaround activities will be divided into several activity characteristic elements, then make sure the priority ordering of the activity characteristic elements according to the all kinds of containment relationships between all two characteristic elements, after that set up optimized objective function, and on this basis, set up fitness function. Then, utilize crossover operator to change the sequence of the logistic activities, and use the mutation operator to figure out the scale of logistical resource, get a optimized activity planning scheme for the turnaround activities. Finally, through an example verify the sequencing optimization method's feasibility and effectiveness [3].

B. Xiao (✉) · S. Ma (✉)
School of Reliability and Systems Engineering, Beihang University, Beijing, China
e-mail: xbp@buaa.edu.cn

S. Ma
e-mail: mashuli1012@126.com

H. Huang · A. Wang
China National Aero-Technology Import and Export Corporation, Beijing, China
e-mail: huanghaiping@catic.cn

A. Wang
e-mail: wangaoqing@catic.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_26

1 Description of Military Aircraft Turnaround Activities

Military aircraft turnaround time is the required preparation time which describes the aircraft from the end of the last mission to return to the task once again dispatched in a continuous mission under the use and maintenance support conditions [4]. The factors which influence turnaround time includes three aspects, the first one is military aircraft own design features, the second one is the tasks type performed by military aircraft, the last one is the support resources, and so on. Turnaround activity sequencing planning is an extremely complex process. It not only effected by the type of task and the support resource consumption, but also by the constraints of logical relationships between any two activities. Therefore, activity sequencing planning is a constrained nonlinear programming problem.

1.1 Description of Support Resources Constraints

Use support resources including crew, varieties of support equipment, facilities, and tool, and so on. The quantity of crew, crew professional and technical levels, support equipment, facilities and tools are available, and the level of support and reliability of the equipment and the equipment degree of match with others all will affect turnaround time [1]. For military aircraft, crews are often divided into four professional, Aeromechanical, Aeronautical Ordnance, Aeronautical Special and Aviation Electronics according to their professional, all professions have their own special work. In the use of job analysis, identify a specific type of task work item of turnaround and the working hours, the logical relationships between any two items, and then considering the characteristics of the ten activities of the relationship between the unit and its influencing factors [5].

There are two types of support activities sorting problems, one based on resource constraints, through reasonable arrangements to find the sequence of turnaround activities; The other is through the activities scheduling to balance variety of support resources and in order to reduce resource reserves, and enhance the likelihood of achieving program, and through balance support resources to reduce the resources idle rate and improve the protection of resource utilization, so as to achieve the purpose of reducing the cost of support activities.

1.2 Turnaround Programme Under Logic Constraints

According to the strict logical relationship between activities, the safeguards activities were determined, and forming a certain logical structure framework. Take air to air combat missions for example, the turnaround activities are expressed as a collection F = {f1, f2, f3, f4, f5, f6, f7, f8, f9, f10}, in it, is the active feature unit its detailed information was described as follows in the Table 1.

Table 1 Air to air turnaround task table

Code	ID	Task item	Professional	Working time (min)	Immediately before the task item
f1	A	Data download	Aviation electronics	2	
f2	B	Outer inspection	Aeronautical special	4	
f3	C	Install drag chute	Aeronautical ordnance	3	
f4	D	Fuel up	Aeromechanical	6	A
f5	E	Loading mission data	Aviation electronics	3	A, B, C
f6	F	Check the system power	Aeronautical special	5	B
f7	G	Hanging missiles	Aeronautical ordnance	7	D, E, F
f8	H	Pilots check	Pilot	3	G
f9	I	Flight parameter self test	Aeronautical special	2	G
f10	J	Work before flying	No	4	H, I

1.3 Turnaround Activities Planning Analysis Under Resource Constraints

While Carrying out turnaround activities plan, resource constraints need to be taken into account, combined support activities sort planning with resource scheduling, achieve the optimal path planning and resource scheduling activities generated simultaneously [6].

At first, divided turnaround activities into several units, and according to the based on the active feature unit determine relationship between the characteristics of cell binding order to priority establish the objective function, and design verification procedures to ensure that the limited nature of chromosomes segment combined. On the basis of the objective function establish fitness function, and design chromosomal crossover and mutation algorithms, and then use genetic algorithm to optimize the activity sorting path. According to the certainty impact factor of turnaround activities to make sure that the relationship between two activity cells, then on the base of not affect the activities' relationship, according to uncertainty empirical data of support resources to establish a safeguards activities planned initial program model [7]. After that, use the mutagenic factor method to adjust the impact factors which are uncertain, and get the turnaround activities planning scheme. And then, make use of fitness to balance the scheme, find the critical path of turnaround activities, choose the best one at last, the turnaround time is the shortest one.

1.4 Relationship Between Two Activity Feature Units

1. Logical order relationship
   The relationship of logical order between support activities is said that the order of activities can't be changed and adjusted. The logical relationship is determined by the design of the aircraft characteristics of the decision, and not allowed to change the logical relationship.
2. Space Interference Relationship
   When support personnel conducting support activities, due to space limitations, a certain safeguards activities carried out by the other one or several safeguards activities spatial interference and influence, and lead to the safeguards activities can't be done with other activities in parallel, so activities come into being before and after order relation.
3. Resource Consumption Relationship
   The configuration of support resources used will affect the turnaround activities time and workflow. In turnaround support activities, different support activities may occupy the same support resources. If the entire support resources are limited, then it may cause a no logical support activities not simultaneously be done. For example, turnaround activities may be affected by the crew constraints, some work can be carried out in parallel, but some work can only be serially. As a confidential work only have a professional staff, while do turnaround activities the Aeronautical Special can only be carried out serial, Aeronautical Special must can do the activity one by one.
4. Security Risks Relationship
   Among turnaround activities, one activity is completed or not will directly affect the other activities of the security or a security risk.

2 Turnaround Activities Sequencing

Considering the support resource constraints when do the turnaround activities plan, these conditions as followed:

1. The logical sequence between the activities;
2. Space interference relationship between support activities;
3. The quantity limits of each sub-activity of the professional support personnel, technical grade;
4. The quantity and types of equipment limitations of protection activity;
5. The relationship between the activities of the security risks (Table 2).

All these activities will serve as the constraints limiting conditions sorting turnaround activities. In order to shorten the turnaround time, when optimizing the support activities these certain sequence must maintain invariability, at the same

Table 2 Logical relationship between turnaround activities table

Activities feature units	Priority activities feature units
f1	Null
f2	Null
f3	Null
f4	f1
f5	f1, f2, f3
f6	f2
f7	f4, f5, f6
f8	f7
f9	f7
f10	f8, f9

time, it consider the multiple allocation of limited resources between projects, so that all projects can be within the prescribed period completed in the shortest possible time.

3 Out Again Based on Genetic Algorithm Optimization Scheduling Preparation Activities

3.1 Gene Encoding

Since genetic manipulation is carried out between the digital, design variables must be optimized volume digitization. The object here is turnaround activities. The usual binary encoding values is suit for the independent variable optimization problems, and the interaction between the independent variables is smaller, but the binary coding process unable to express scheduling problem, so using natural number coding method there [8].

A combination of chromosome fragments which correspond to turnaround activities of one particular task are series of activities, which stand for the specific task of turnaround activities. If there are n items in a task, then there are also n-chromosome combination or n genes chromosome fragments that can be expressed as {G1, G2, …, Gi, …, Gn}, Gi represents the i gene which to be done. Each gene consists of three parts, one part is the active feature unit type code, one is Professional Code, and the other one is the time of every activity. In it, A, B, …, F… denote the active cell f1, f2, …, fn; Aeromechanical, Aeronautical Ordnance, Aeronautical Special and Aviation Electronics use the Figs. 1, 2, and 4 to indicate, and the pilot with five, at last no professional work required optional arbitrary numbers with zeros to represent. So where (A,1,2) represents the data download f1, avionics professionals need to complete the, and the work needs 2 min. As we all know, the air to air turnaround activities of genes encoding feature unit as shown below (Table 3).

Table 3 Turnaround activities feature units gene encoding

Activities code	Gene encoding
f1	(A,1,2)
f2	(B,2,4)
f3	(C,3,3)
f4	(D,4,6)
f5	(E,1,3)
f6	(F,2,5)
f7	(G,3,7)
f8	(H,5,3)
f9	(I,2,2)
f10	(J,0,4)

3.2 Generation of the Initial Population

According to the gene coding rules and length of the chromosome containing N randomly generated initial population of chromosomes as the initial solution of the problem. Initial population size should be appropriate to the scale of N. If N is too small, it easy to fall into local optimal solution; if N value is too large, it will be lower operating efficiency. In practical application experience or experimentally, it only be determined in accordance with generally recommended ranges from 20 to 100.

Because turnaround activities will be subject to many conditions, so randomly generated chromosomes may be invalid, in order to ensure the initial population all validity, the algorithm to generate chromosomes as shown in Fig. 1. Turnaround activities randomly generated two valid chromosomes are described as Figs. 2 and 3.

3.3 Population Calibration Procedure

To ensure the implementation of individual genetic algorithm effectiveness, it must be determined processing priority verify the validity of the relationship of the individual according to the mandatory constraint [9]. The method of calibration starts from the last gene of the chromosome, traverse the entire chromosome, and determine whether genes individually meet the constraints. The algorithm was shown in Fig. 4.

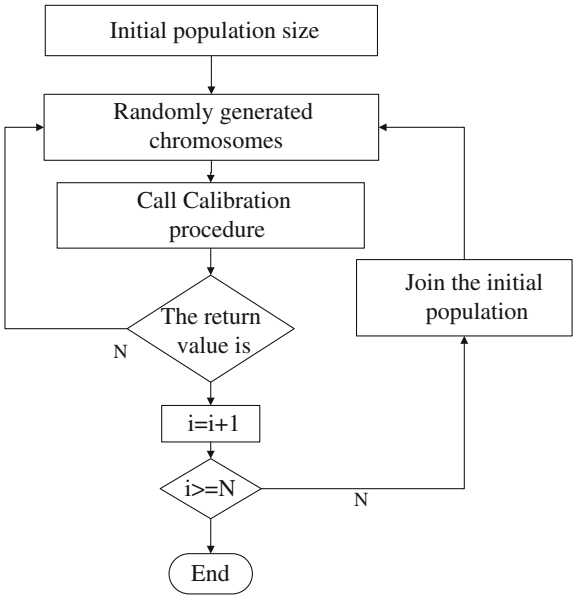


Fig. 1 Generation flow chart of initial population

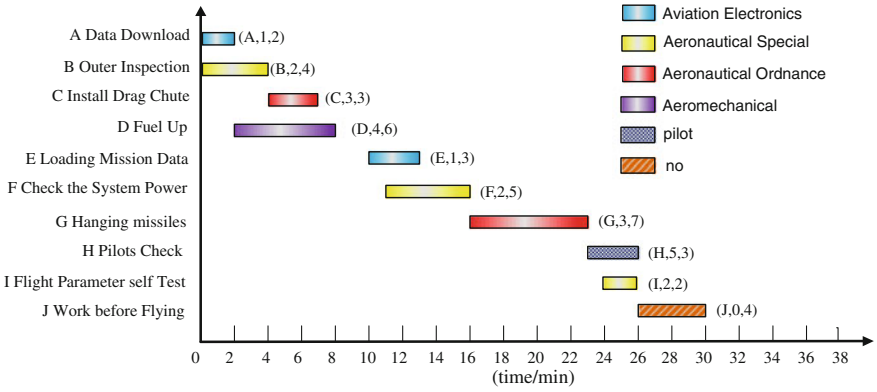


Fig. 2 Turnaround activities effective program—a combination of chromosome fragments

3.4 Fitness Function

Turnaround activities program is a multi-objective constrained optimization problem. Take turnaround time, support resource as campaign optimization goals. According to the importance of these two goals, and then multiplied by the weight coefficient, after that summation. This put a multi-objective optimization problem is transformed into a single objective optimization problem. This article will deal with the mandatory constraints alone, in the verification procedures to ensure the

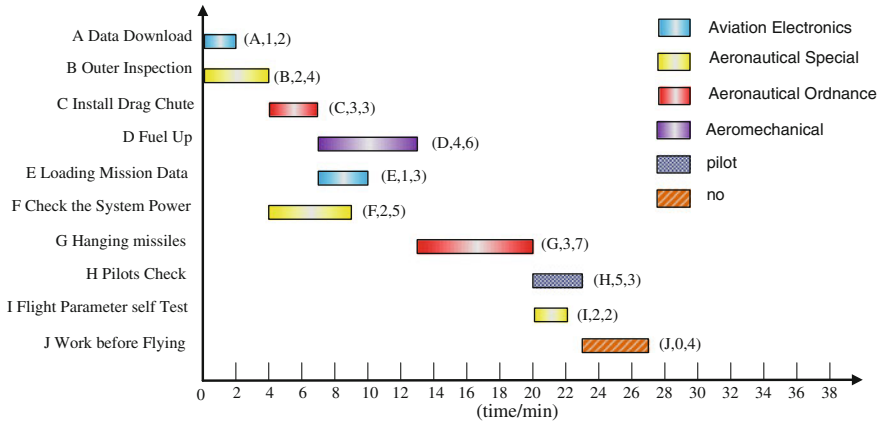


Fig. 3 Turnaround activities effective program—b combination of chromosome fragments

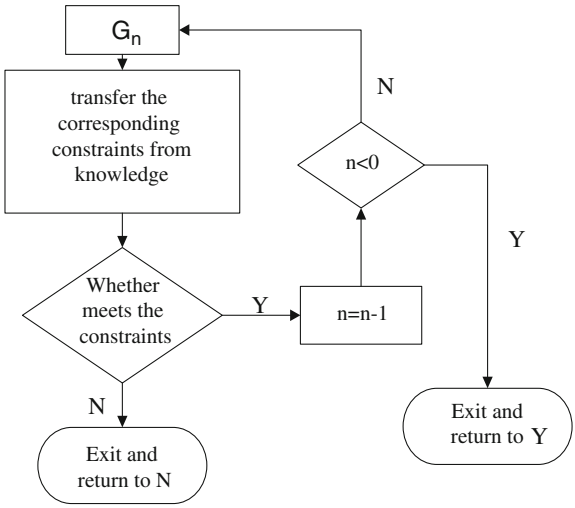


Fig. 4 Chromosome checking program

population of each chromosome is legitimate, so the objective function can no longer consider the constraints; in this way, the original problem eventually became an unconstrained single-objective optimization problem [10]. It is assumed that a combination of n-dimensional space of chromosomal segments, the so-called n-dimensional chromosome segment refers to a combination of each chromosome consists of n genes. Then, the objective function can be expressed as:

$$\min C(x)(x \in R^n) \tag{1}$$

$$C(x) = a_1 C_m(x) + a_2 C_r(x) \tag{2}$$

$$C_m(x) = \sum_{i=1}^{i=1\,n-1} \max[\delta(G_i T)] \tag{3}$$

$$C_r(x) = \sum_{i=1}^{i=1\,n-1} [\delta(G_i Q)] \tag{4}$$

In this formula:

X	Chromosome fragment combinations;
a_1, a_2	Turnaround time and the weight coefficient of the scale of support resources and the protection. Support resources scale of the number of professionals and devices which are occupied by activities. According to weigh standards to determine the weighting factor, set $a_1 = 0.9$, $a_2 = 0.1$;
$C_m(x), C_r(x)$	Turnaround activities time, the scale of support resources;
$G_i T, G_i Q$	The spends time of Gene Gi, the scale of Gi support resources;

In genetic algorithm, the merit of this measure is to fitness. According to the size of fitness, decide that whether the certain individuals are breeding or die. The fitness greater, the individuals more gifted. The fitness function transforms based on the objective function, its function value should remain positive, the objective function value of the function correspondence, therefore there have the same extreme points, the fitness function F (x) can be expressed as followed:

$$F(x) = \begin{cases} \frac{1}{C_{max}}(Cx < C_{max}) \\ 0(\text{others}) \end{cases} \tag{5}$$

In this formula: C_{max} is a sufficiently large constant, generally the $C(x)$designed as the maximum value. As in the above two figures, support activities in the basic resources used there is no competition between professional requirements and no conflict between competition. Support resources can be quantified as an economic indicator, which uses a variety of support resources required to describe the cost of performing this activity support resources scale.

In the air-to-air turnaround activities, the cost of support equipment and personnel needs set 7,000 RMB.

After calculation: Program A fitness is: 0.036, Program B fitness is: 0.040.

3.5 Mutation

Mutation is one method that generating new individual mutations. The algorithm described as follows:

1. Randomly select a chromosome from the current population.

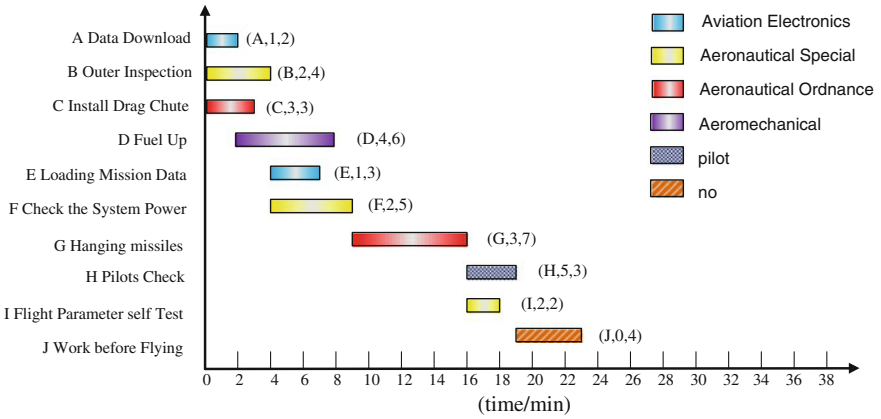


Fig. 5 Turnaround activities effective program—c combination of chromosome fragments

2. Still set the number of chromosome genes is n, n can randomly generated positive integer as the adjustment point from 1 to n.
3. Within the permissible range of genetic point adjustment, forming a new individual.
4. Calls checksum algorithm to determine the effectiveness of a new individual, if the generation group is effective, added to the next generation; otherwise return to step (2).

3.6 Termination Method

There are two the most common method of termination. First, the number of iterations reached regulation, the calculation terminated; the second one is the result of several iterations of optimal individual fitness does not change or changes very small, calculated terminated. The resulting optimal activity sequencing program as follows Fig. 5.

After calculation, the fitness is: 0.047.

4 Conclusion

Turnaround activities feature units division and its work project flow generation, constraints analysis, the objective function and establish the fitness function check, copy, crossover and mutation algorithms is the key in this article. In this paper, through verify and calculate the air to air turnaround activities, getting more satisfactory results on the genetic algorithm, but in air to air turnaround activities, the

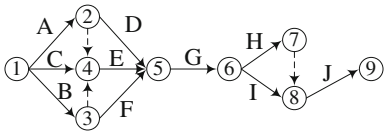


Fig. 6 Air to air turnaround activities Project Coordinator Figure

resource conflicts are not obvious, so the calculation in this article is relatively simple.

According to the Fig. 5, drawing the air to air turnaround activities Project Coordinator, as follows (Fig. 6).

The key work items of air to air turnaround activities are:

$$B \rightarrow F \rightarrow G \rightarrow H \rightarrow J$$

The turnaround time is 23 min. However, the division of turnaround activity needed to identify the characteristics information of activities, such as personnel professional, personnel technical grade, the type of support equipment, and so on. But these characteristics in some tasks reflected not obvious, however, this article uses the method that using one parameter instead of the all support process cost to describe the features.

References

1. Wei T (2009) Research on optimization of multi-project network planning based on genetic algorithm [D]. China University of Geosciences, Beijing, pp 30–33
2. Fu Z (2012) Multi-resource balance optimal decision-making based on genetic algorithm. Equipment Manufacturing Technology, Guangxi
3. Yi W, Rui K, Hailong C (2008) Calculation method for military aircraft's turnaround time. J Beijing Univ Aeronaut Astronaut 34(12):1415–1418 (China)
4. Liu S, Wang M, Tang J (2001) The optimization algorithms for solving resource-constrained project scheduling problem. Control Decis 34(12) (Shenyang, Liaoning, China)
5. Zhao J, Zhu H, Wang W (2010) Operations research on technical support for aviation material, 1st edn. National Defence Industry Press, China, pp 126–149
6. Xiao Y (2000) Processes and step sequencing strategy based on hierarchical planning. Mech Electr Eng Mag 17(4) (Hunan, China)
7. Jerome B, Jean C (1990) A Monte-Carlo algorithm for path planning with many degrees of freedom. Computer Science Department, Stanford University, USA
8. Alluru GK, KM R (2009) Optimization of operations sequence in CAPP using an ant colony algorithm. Int J Adv Manuf Technol 29:159–164
9. Rudolph G (1994) Convergence analysis of canonical genetic algorithms. IEEE Trans Neural Networks 5:96–101
10. Zheng XM, Thomas S (2004) Applying a genetic algorithm-based multi-objective approach for time-cost optimization. J Constr Eng Manag ASCE 130:168–176

Business Intelligence and Service Oriented Architecture—Improving IT Investments

Indira Venkatraman and Paul T. Shantapriyan

Abstract Information Technology does not solve business problems. Rather, IT can be used as an enabler to meet goals, targets and position the organization in the minds of the customer. A Service Oriented Architecture (SOA) can enable higher level services from more primitive business processes. The result is a flexible, agile organization. However, in some cases, without a business model, an SOA can be implemented but the result is islands of incompatible services. In the era of data analytics, the evolution of business intelligence has assumed a more dominant role. Proponents argue that Business Intelligence subsumes SOA. Other practitioners argue that Business Intelligence (BI) and SOA are distinct from each other as each aims to deliver different perspectives to an organization. This chapter challenges this view by developing a framework where both SOA and BI are pivotal to delivering the business model. The analytics, business intelligence and data mining handshake with the SOA, allowing decision makers to plan, coordinate and control resources to meet the goals, targets and performance measures set for the organization. Business Intelligence and SOA need not operate on different levels for an organisation. In an era where companies are trying to go from barely surviving IT expenses towards developing an IT portfolio, the more options for technologies supporting asset data management, warehousing and mining to interact and co-exist are necessary. Putting aside arguments of Business Intelligence versus SOA, we propose that they can handshake and bring out the collaborative synergies to enable organizational decision making to address the competitive challenges in today's global marketplace.

Keywords Decision support and optimization methods and tools · Business intelligence · SOA

I. Venkatraman
Principal Consultant, Intelligent Capital, Hobart, Australia
e-mail: indira@intelligentcapital.info

P.T. Shantapriyan (✉)
Tasmanian School of Business & Economics, University of Tasmania, Hobart, Australia
e-mail: Paul.shantapriyan@utas.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_27

1 Introduction

Ruben finished his picture with a "future" diagram."...Technology-wise, we are moving more towards SOA, more real time operations, and expanding database structures more into data warehouses that we can mine for trends and other BI." [1].

When an organization looks to the future and wants to go from barely paying bills to investing into portfolios and infrastructure, the issue is to identify what works, what matters; and not base decision making on fears, or be constrained by phobias of forbidden combinations of technology held rival schools of IT thinkers. IT is an Enabler; not a magic bullet to make decisions automatically for a strategic position in the mind of the customer. BI is a process by which knowledge is created, captured, shared and leveraged for a company to compete and succeed [2]. A typical business intelligence system include functions such as reporting, multi-dimensional analysis, querying tools, online analytical processing tools, forecasting, data-mining, and advanced visualization capabilities [3]. Whereas service-oriented architecture (SOA) is a group of well-defined services which can be combined, reused and communicated with each other over networks [4]. SOA is an architectural design to enable flexibility from loosely coupled systems that are aligned to the organization [5]. At a glance, SOA and BI exhibit differences. A quick overview of the general differences is shown in Table 1.

BI and SOA have conflicting needs and principles. BI needs data that is hidden within an organisation and tries to get to it by Extracting Transforming and Loading data (ETL) [6]. The process of ETL might be easy with SOA as within SOA environments. One of the outstanding advantages is individual services that can be accessed without knowing the underlying system platform [7]. SOA with access to a database/warehouse and entries of data within such environment tend to have metadata (data on data) forms a solid environment from which Business Intelligence staff extract what data they need to process into decision making information.

Proponents of SOA might argue that most Service Oriented Architectures with in-built or add on BAM abilities (Business Activity Monitoring) might do away with the need for anything BI. Supporters of BI on the other hand might (and rightfully) say the BI has been in existence prior to dominance of Information Systems in organisations.

Table 1 Differences between service oriented architecture and business intelligence [5]

SOA	BI
Enterprise-oriented	Subject-oriented
Middleware	Extracting transforming and loading (ETLs) and presentation
Transactional	Non-volatile
Reliant on web services and message-level processing of data	Reliant on very large data bases (VLDBs) and terabytes of data
Real time	Scheduled

The power of IT to enable business intelligence (trend analysis, forecasting, modelling, etc.) grew exponentially [8]. The phrase "Big Data" to identify that business intelligence is being run by "data scientists" [9]. For the purposes of this chapter BI is used to reflect the nuances captured by big data. BI can be likened to information over load. For example in an SAP environment, there can be as many as 5,000 cost pools (aggregation of cost into one area of analysis). Each cost pool could have dozens of cost drivers (activity that drive costs). The ability to model such cost behaviour is time consuming. The question that remains is whether we know which if these cost pools is value added.

To understand value, lot of organisations have looked at cost reduction as equal to value generation. However value lies in the eyes of the customer. Eliminating a set of costs (business processes) may be eliminating value in the eyes of the customer. For example a water board in the UK decided to outsource a call centre to an Asian country [10]. Unfortunately the water board clients were based in Yorkshire. The accent of the clients was not understood by the call centre operators. The clients were furious. They shifted water supplier. The water board re-established a domestic call centre. The lesson learnt was that BI however profoundly analytical within an organisation perspective forgot to address the customer centred issues.

This customer centric approach was one of the assertions for SOA. Proponents of SOA have used service science/ service dominant logic as one of the core themes that drive SOA.

We are going to argue that neither SOA nor BI has to supersede, replace or include the other. They can exist as they are in an organisation and can handshake at various intervals to create better value, and better learning of their interaction without lose to both parties.

2 A Few New Ideas

2.1 Hidden Data

Basically, for us to achieve BI Nirvana, all we need is "just" one input: data. BI needs the data that is hidden within the organization's systems [11].

Extracting, transforming and loading data to transform into valuable information for decision making will need to meet the simplest criterion: usefulness to the decision maker. The opinions as to whether or not a packet of data is useful may vary from BU to BU and person to person. Data within the organisation can be hiding in processes, minds of employees, managers or customers. SOA aims to find these packets of data to enable decision making. The architecture for this service orientation requires investment and resources. The contracts for the data warehouses, various systems embedded within the master system. SOA tries to form a comprehensive fabric of computers and services that work in an integrated manner. In reality, loosely coupled systems link together islands of data in a variety of places. IF SOA dominates BI, there is a reduction in redundancy of data as well as

better integration and coordination of resources. However, BI should not be constrained by the architecture of the current IT.

One word of caution is that such separation means the need to more agile in activity and process analysis and lateral thinking of the big data scientists. An ideal business environment is one where there is no hidden data and there is complete transparency. But not all businesses can behave that way.

2.2 Meta Data

Simply put this is data about data. The allocated key, value pairs for example in a database are a form of metadata. They tell you what would be otherwise a pile of jargon, actually is. Another example would be fields in form design that help one note the name, middle name, and surname. In some cases metadata could be the hidden data. It is also helpful to quickly assess patterns and flow charts. However, metadata describe the simple data and what is needed is the ability to transform the sets of data to information to address the knowledge needs of the decision maker. We now explore the concept of knowledge, in particular tacit knowledge.

2.3 Tacit Knowledge

Tacit knowledge is valuable insights drawn from and highly subjective intuitions carried by people in their minds [12]. This tacit knowledge is hard to codify [13]. For example, in Japan, the socialisation between employees in the project team for developing a bread making machine is important. The bread was not soft enough [13]. An employee then observed a master baker and she noticed that the master baker twisted the dough, the key stage for softness. By observing the tacit knowledge, and sharing that process with the machine development team, the tacit became explicit [14], allowing codification. In this way, the hidden data now becomes visible.

This example of tacit knowledge underscores two important points: the perception or cognition in the minds of the individual and the interaction between people to develop knowledge. BI has relied on the skills of data scientists and assumes the knowledge generation process. However, the importance of observation and cognition is pivotal for BI to work. SOA has however, looked at the architecture or software protocols to generate knowledge. The cognition of people and the socialization and interaction needed in the knowledge transformation is unserved by SOA. Rather, knowledge is held to be context free, objective in traditional SOA proponents. However, human decision making can be context driven as well as be pluralistic or hedonistic when engaging with IT.

2.4 Human Beings

Plurality of views is not a constraint to knowledge or business intelligence. For example, when the internet service on cell phones was introduced in Japan, the traditional mind set was an enhancement of existing services [14]. However, outsiders to this industry held the view that young people had fun. This synthesis of competing views allowed the telecommunication company to move from the utilitarian views of business people to the hedonistic views of the younger generations.

People and their process of knowledge generation can leverage the SOA architecture to allow networks within and between organizations. SOA promises reliability, scalability, integration, reducing data replication as well as building networks between people and organizations [15]. Such aims are truly noble where the disruptive technologies such as cloud allowed data marts (such as Ebay) to reduce data held in fifty different folders to be substantially reduced [16].

However, if the design of the SOA architecture fails to deliver on time, does that mean that the BI analytics and decision making is held up? Is the cost of such failure (or lack of success) of the architecture reduce the importance of decision making of BI? By allowing BI to be a pillar of an organization portfolio as well as SOA as another pillar of the IT portfolio, people can socialize with each other and engage with the technology to facilitate improved decision making, analytics as well as knowledge. In this manner, the architecture and the knowledge creating processes of the employees, managers can handshake with each other. We propose that they can handshake and bring out the collaborative synergies to enable organizational decision making to address the competitive challenges in today's global marketplace.

3 Let BI and SOA Handshake

This chapter argues that letting BI and SOA be themselves in terms of entity identity within an organisation and where need handshake to produce desired results. Thus the hidden data with in BI structures and meta data within SOA structures do not get modified and hence altered for future reference. This handshaking also might make sure that tacit knowledge is extracted with offending anyone or altering anything. The following is a diagrammatic representation of an environment where BI and SOA handshake. It is an enterprise view, and it is agile. Components of SOA and BI are not constricted within this framework (Fig. 1).

This is only a working model. But it is clear enough to give the right picture about handshaking. The elements of process of Business Intelligence are intact within the enterprise from where data is collected all the way to where it becomes information that can be processed for decision making and understanding.

Simultaneously SOA has its own existence with data, meta data, customer architecture and architecture contracts. These subsets of SOA bring value to BI that

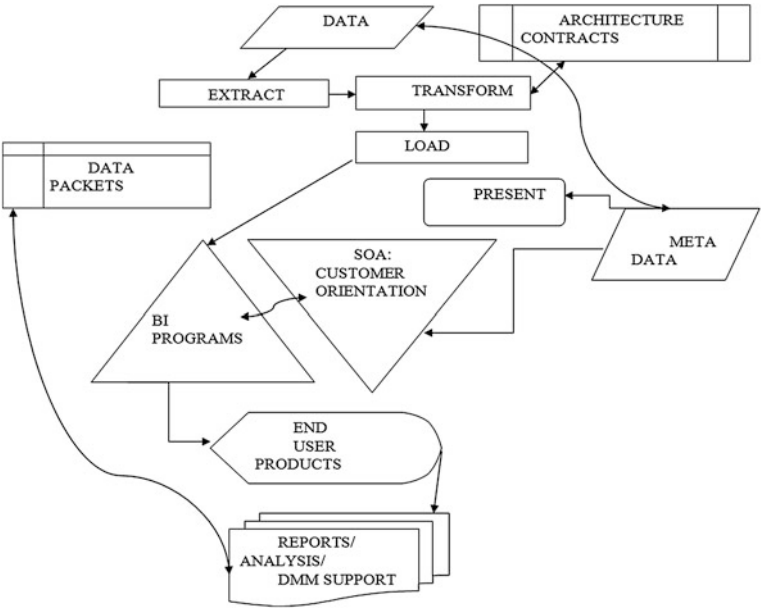


Fig. 1 Handshaking of BI and SOA; agile, enterprise view

otherwise ends up being hierarchical and fixed in its views. Parallel importance should be given in viewing how at need intersections BI information can be re processed as data packets for SOA and possibly gain new meta data. And hence enhance customer orientation by added information.

4 Conclusion

The agile enterprise view considers people (managers, employees and project teams) as pivotal in the utilization of BI and SOA. The BI programs (be it software or teams of people) can engage with the data (after ETL) while ensuring that the ensuing architecture is customer centric. This customer orientation is central to the handshaking between SOA and BI. BI evaluates models and analysis opportunities in the customer and market space for new products and services. The SOA is designed and continuously adapts to meet these shifting landscapes of services, products and opportunities. The ensuing outputs (reports) inform BI for further iterative handshaking. This real time interaction between BI and SOA allows both pillars of IT to work in conjunction, rather than be subsumed as subsets of the other. This loose coupling does place greater burdens on SOA but the tacit knowledge generated by potential BI is worth stretching SOA, both from a design perspective as well as challenges to implementing and aligning the architecture.

References

1. Austin RD, Nolan RL, O'Donnell S (2009) The adventures of an IT leader. Boston Harvard Business School Press, Boston, pp 64–65
2. Microstrategy, 2007 as cited in Chan LK, Sim YW, Yeoh W (2011) A SOA driven business intelligence architecture communications of the IBIMA. http://www.ibimapublishing.com/journals/CIBIMA/cibima.html
3. Earl 2005 as cited in Chan LK, Sim YW, Yeoh W (2011) A SOA driven business intelligence architecture communications of the IBIMA. http://www.ibimapublishing.com/journals/CIBIMA/cibima.html
4. Choi JC, Nazareth DL, Jain H (2010) Implementing service oriented architecture in organizations. J Manag Inf Syst 26(4):253–286
5. Wik P (2011) Service-oriented architecture and business intelligence. Serv Technol Mag, August, pp 1–17
6. Aziz S (2006) Service oriented architecture: look beyond the myths to succeed. Retrieved from InfoSys: http://www.infosys.com/services/systemintegration/white-papers/SOA-look-beyond-myths-to-succeed.pdf
7. Kodali (2005) as cited in Chan LK, Sim YW, Yeoh W (2011) A SOA driven business intelligence architecture communications of the IBIMA. http://www.ibimapublishing.com/journals/CIBIMA/cibima.html
8. Davenport TH, Harris JG (2007) Competing on analytics: the new science of winning. Harvard Business School Press, Boston, p 240
9. Sharma R, Dijaw V (2011) Realising the strategic impact of business intelligence tools. J Info Knowl Manag Syst 41:113–131
10. Stringer C, Shantapriyan P (2012) Setting targets. Business Expert Press, New York
11. Rotem-Gal-Oz A (2007) What is SOA? Downloaded from http://www.rgoarchitects.com/files/SOADefined.pdf
12. Nonaka I (2007) The knowledge-creating company. Harv Bus Rev July August 85:162–171
13. Gourley S (2006) Conceptualizing knowledge creation: a critique of Nonaka's theory. J Manag Stud 43(7):1415–1436
14. Nonaka I, Toyama R (2007) The theory of the knowledge creating firm: subjectivity, objectivity and synthesis. Ind Corp Change 4:419–436
15. Davenport T, Patil D (2012) Data scientist: the sexiest job of the 21st century. Harv Bus Rev 1–8
16. Davenport T, Barth P, Bean R (2012) How big data is different. Sloan Manag Rev 43–46

Effective Guided Wave Technique for Performing Non-destructive Inspection on Steel Wire Ropes that Hoist Elevators

Peter W. Tse and J.M. Chen

Abstract The steel wire ropes that hoist elevators will eventually become rusted and then cracked as they always expose to air that contains high salinity content. A number of rope broken accidents occurred because of inappropriate maintenance that causes the ignorance of broken wires in a steel rope. Elevator ropes have very complicated structure as they are formed by twisted many small wires together to form a bigger rope. Such complicated structure creates difficulty to currently available non-destructive evaluation methods. To avoid the accidents, proper maintenance equipped with effective non-destructive testing method that must be presented. However, as of today, most of the elevator rope inspections are depends on human visual inspection. Such inspection fails to detect faults occurred in internal wires and even external wires due to the covering of grease on the ropes. The aim of this study is to introduce a new and effective technique based on ultrasonic guided waves for inspecting the faults occurred in internal and external wires. The experimental results demonstrate that the ultrasonic guided wave can detect the locations of broken wires and then determine the number of broken wires effectively. The experimental results also show that the PZT sensor is more suitable for working as a transmitter/emitter and the MsS is better working as a receiver. For grease covered ropes, the results show that the range of inspection distance will be decreased due to the attenuation of the reflected signal energy. Nonetheless, the reflected signal caused by the defects can still be clearly observed even the rope was covered by grease. Although the current results are promising, more sophisticated method must be developed so that the proposed technique can be applied to on-site ropes tests. The future research may include the development of optimizing the operation parameters of guided wave and the design of practical sensor for operating ropes.

P.W. Tse (✉) · J.M. Chen
The Smart Engineering Asset Management Laboratory (SEAM), Department of Systems Engineering and Engineering Management (SEEM), City University of Hong Kong, Tat Chee Ave, Kowloon, Hong Kong, People's Republic of China
e-mail: Peter.W.Tse@cityu.edu.hk

J.M. Chen
e-mail: jingmchen2-c@my.cityu.edu.hk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_28

Keywords Ultrasonic guided wave · Non-destructive evaluation · Condition monitoring · Rope and cable inspection

1 Introduction

The structure of Steel wire ropes/cables is complex, as they are formed by many twisted and inter-locked wires. A tested hoist wire rope (8 × 19S + FC) has a fibre core which is shown in Fig. 1. The wire ropes are widely used in elevators for hoisting or supporting heavy objects. An elevator bears heavy loads during daily operation. Day after day, loading and unloading heavy objects could make steel wire ropes/cables prone to severe defect. A saline and humid environment in particulate country areas escalates the wear on ropes and cables, which can lead to their sudden breakage. For instance, rope accidents have caused hoist lifts to sudden plunge to the ground, resulting in human casualties. To avoid that, several methods in existence today are applied for wire ropes/cables inspection. These methods include the use of magnetic flux leakage testing [1], human visual testing, and magnetostrictive testing [2].

The principle of the magnetic flux leakage testing method is to inspect wire ropes by using a high strength magnetic field that is moving along the wire ropes. If there are any defects occurred in the specimen, the leakage of the magnetic field will be detected by the transducer. Human visual testing method is the traditional rope detection technologies. It is mostly dependent on visual inspection by measuring rope diameter. The reduction in diameter illustrates corrosion, breakage and failure in a wire rope. Magnetostrictive testing is a method using magnetostrictive techniques. Magnetostrictive effect is caused by the permanent magnet circuit

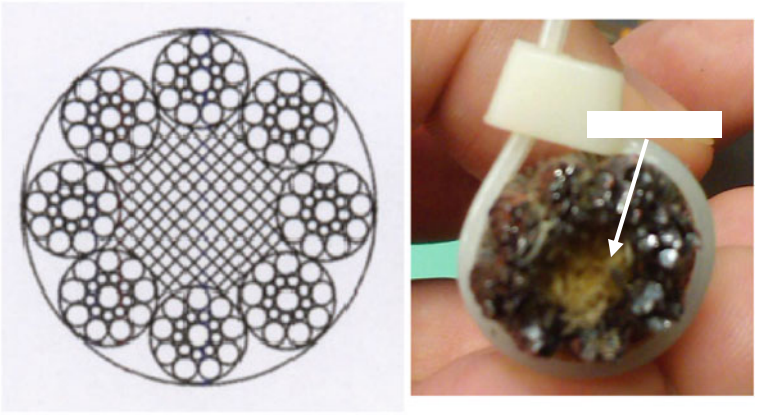


Fig. 1 Cross-section view of a hoist wire rope (8 × 19S + FC)

provided by the state of strain of a ferromagnetic and the dynamic magnetic field provided by hard sensing coil. The magnetostrictive effect will generate ultrasonic guided waves, which are elastic waves propagating in measured objective. The ultrasonic guided wave has advantages of high sensitivity, long-distance inspection with low attenuation, and line to line inspection. In addition to magnetostrictive testing method, a piezoelectric testing method based on guided wave technology is proposed for wire ropes inspection. The piezoelectric transducers (PZT) are made of length expander-type piezoelectric materials and distributed axisymmetrically, which ensures that proper modes are excited while generating guided waves [3], and this technology has widely been used in pipe inspection at present.

For all the abovementioned technologies, they have their own advantages as well as weaknesses. The weaknesses include difficulties in sensor installation, complications in analysing the reflected wave signals caused by defects occurred in the inspected ropes, and inconsistence in the results, etc. Nowadays, the techniques using ultrasonic guided wave are considered popular non-destructive testing (NDT) methods due to their long-range inspection abilities and high degree of sensitivity in detecting defects. The present chapter will focus on assessing the effectiveness of an ultrasonic guided wave technique on inspecting steel wire ropes that hoist elevators.

2 The Guided Wave Transduction System

2.1 The Types of Available Sensors

Although the ultrasonic guided waves have been applied for wire rope inspection, the design of an optimal sensor configuration is necessary in order to excite required wave modes that will minimize the complication embedded in the received guided wave reflected by different types of defects. As mentioned earlier, ultrasonic guided waves can be generated by magnetostrictive testing and piezoelectric testing methods. As shown in Fig. 2, two sensor configurations, namely the Piezoelectric Transducer (PZT) and the Magnetostrictive Sensor (MsS), were chosen as the sensors for emitting and receiving the guided waves when inspecting ropes. Two wave based measurements were applied in this experiment. They were pulse-echo and pulse-catch modes, respectively. In the pulse-echo mode, signals reflected from the defect edge were excited and received by the same transducer. Different from the pulse-echo mode, in the pulse-catch mode, signals were excited by the transmitter and received by the receiver.

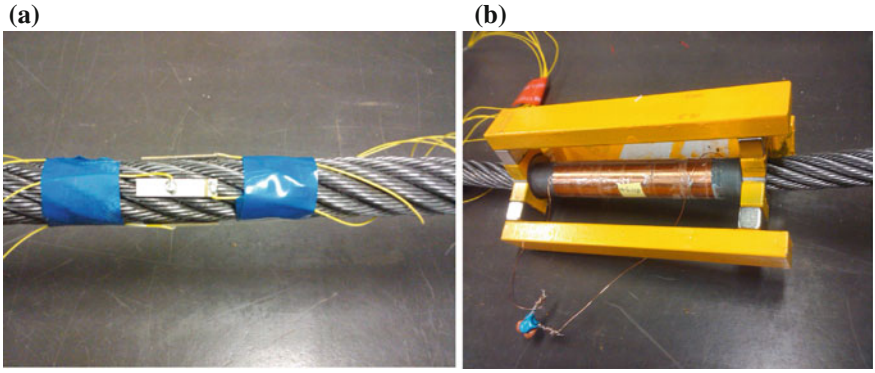


Fig. 2 The two sensor configurations used for inspecting ropes. **a** The PZT sensor. **b** The MsS sensor

2.2 *The PZT Sensors Installed for Working in the Pulse-Echo Mode*

The installation location of the PZT sensor in a 2.02 m wire rope is shown in Fig. 3. Based on our research findings [4], the PZT was working in a pulse-echo mode and the working frequency was set at 100 kHz. The gain for the receiver signal was about 40 dB. The peak amplitude of the "Echo from right end" was about 1 V. Meanwhile, the signal reflected from the left end (propagated about 2.88 m) and the echoes from both ends (propagated about 4.04 m) were received by the PZT. Figure 4 depicts the received reflection signal waveform by PZT sensor.

In Fig. 4, it shows one of the advantages of using the PZT is its high energy conversion efficiency. However, the PZT sensor must be pasted or pressed onto the wire rope surface, which restricts its application for real wire rope detection. Moreover, when the PZT sensor was working at the pulse-echo mode, the oscillation signals deduced by the wafer itself was overlapping with the received signal waveform especially at the beginning of the reflected signal waveform as shown in Fig. 4. Therefore, PZT sensor is more suitable for working as a transmitter/emitter, rather than as a receiver.

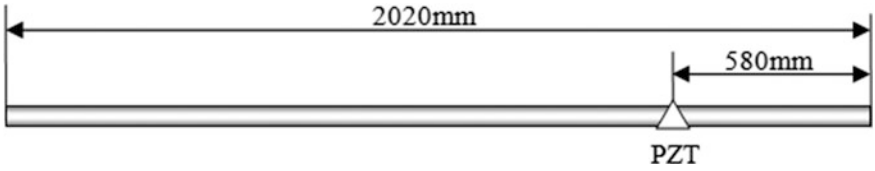


Fig. 3 The installation location of the PZT in the rope

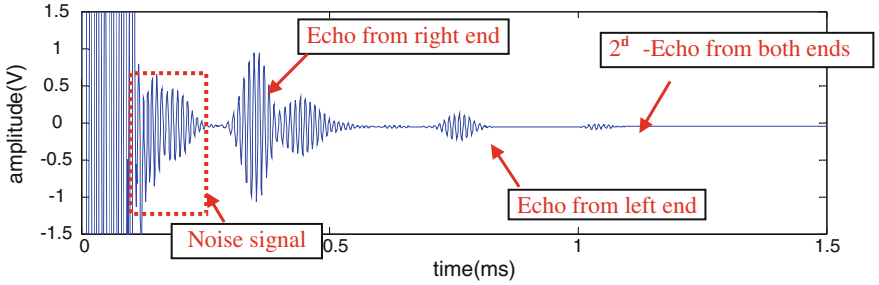


Fig. 4 The reflected signal waveform received by the PZT sensor

2.3 The MsS Installed for Working in the Pulse-Echo Mode

The installation location of an MsS in a 2.02 m wire rope is shown in Fig. 5. Figure 6 depicts the received signal waveform when the MsS was working at 100 kHz in the pulse-echo mode. Although the received signal was amplified by 60 dB, the signal reflected from the left end of the wire rope still had a very low amplitude (<0.1 V). The signal reflected from the left end could not be detected by the MsS. Hence, it is obvious that the PZT had better performance in emitting guided waves by comparing the waveforms as shown in Figs. 4 and 6.

As shown in Fig. 6, the MsS is capable of conducting non-touching inspection to the wire rope. However, it has very poor performance in emitting the desired guided

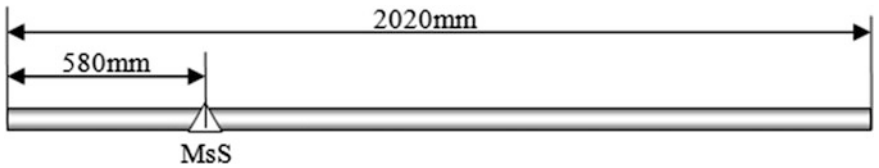


Fig. 5 The installation location of the MsS in the rope

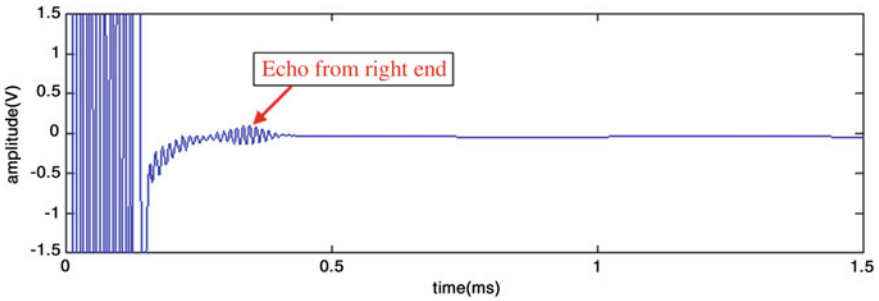


Fig. 6 The reflected signal waveform received by the MsS

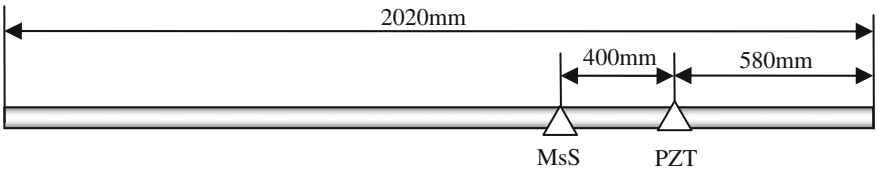


Fig. 7 The installation locations of the PZT and MsS in the rope

waves. Based on the result of previous experiments, when the MsS was working at around 100 kHz, its detection depth was less than 1 mm due to the skin effect of the electromagnetic field. That is, the MsS is more preferable to work as a receiver to receive the guided waves reflected by different types of defects.

2.4 The PZT Works as a Transmitter and MsS Works as a Receiver

In another experiment, the PZT sensor worked as a transmitter/emitter of guided wave and the MsS worked as a receiver of the reflected guided wave. The installation locations of the PZT and MsS are shown in Fig. 7. The gain for the receiver signal was set at 40 dB, and the peak amplitude of the "direct transmission wave (DTW)" was about 1 volt. This peak amplitude is in the same magnitude as that showed in Fig. 2. The "echo from right end" and the "echo from left end" (propagated about 2.48 m) were detected by the MsS. Although the amplitude of the "echo from right end" was smaller than that showed in Fig. 4, it could still be clearly identified. Because the PZT and the MsS was working as a pulse-catch mode, the oscillation signals (noise signal) deduced by the sensor itself was minimized, so that each wave packet in Fig. 8 could be distinguished and tagged with different reflections caused by different defects or conditions of the inspected rope.

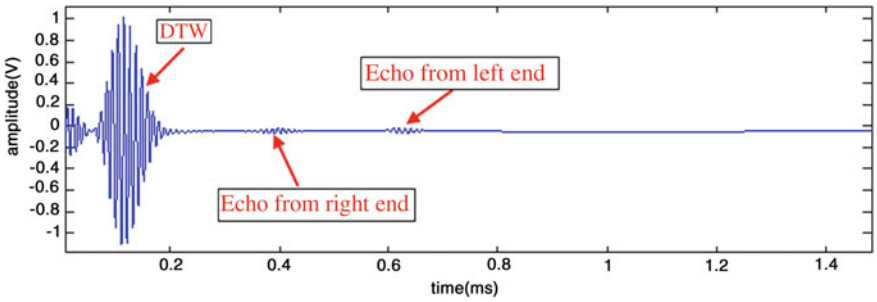


Fig. 8 The reflected signal waveform received by MsS when the PZT was working as the transmitter

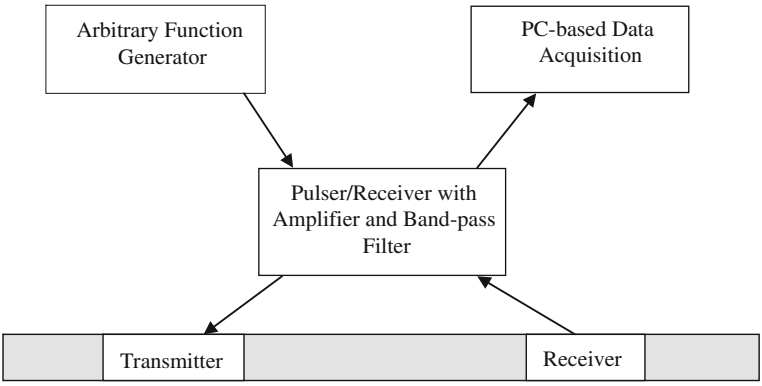


Fig. 9 The experimental set-ups for steel ropes/cables defect inspection

3 Experimental Set-Up and the Results on Broken Wire Detection

The experimental setup for the tested steel wire rope/cable is shown in Fig. 9. Experimental set-up for broken wire detection consists of arbitrary function generator, pulser/receiver with amplifier and band-pass filter, two types of transducers and data acquisition based on PC From the abovementioned experiment, two types of sensors, the MsS and the PZT were used for determining the number of broken wires. Again, the PZT was the transmitter and the MsS was the receiver. The whole transduction system worked in a pulse-catch mode.

3.1 The Determination of Number of Broken Wires in a Steel Wire Rope

The installation locations of PZT and MsS were shown in Fig. 10. The emitted L (0, 1) mode guided wave propagated in both sides of wire rope. The propagation paths of guided wave in the rope are illustrated in Fig. 10. The direct transmission wave from the PZT to MsS, marked as "DTW", was firstly received by the MsS. The wave packet of "DTW" was overlapped with the initial pulse wave due to the MsS receiver was close to the PZT transmitter. The echo wave from right end of the rope, which is marked as "Echo from right end", was received by the MsS after the propagating path reached 1.36 m. The echo wave from left end of the rope is marked as "Echo from left end".

Figure 11a shows the waveform of a signal received from a healthy hoist wire rope. The waveform can be easily identified because there is no overlapping signal or trailing signals occurred as the MsS was the receiver. Figure 11b shows the received signal waveform from a defective hoist wire rope.

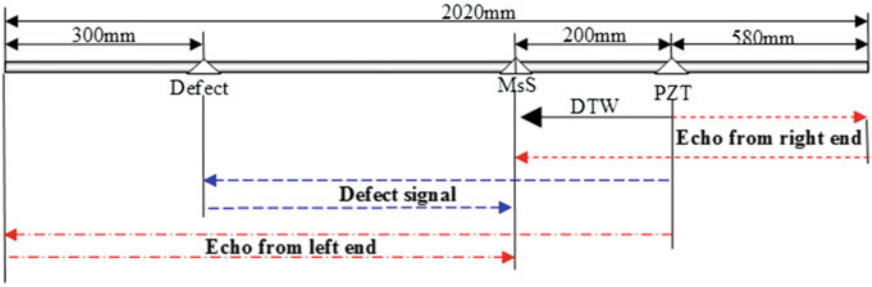


Fig. 10 Locations of PZTW and MsS onto the wire rope and propagation paths of guided wave

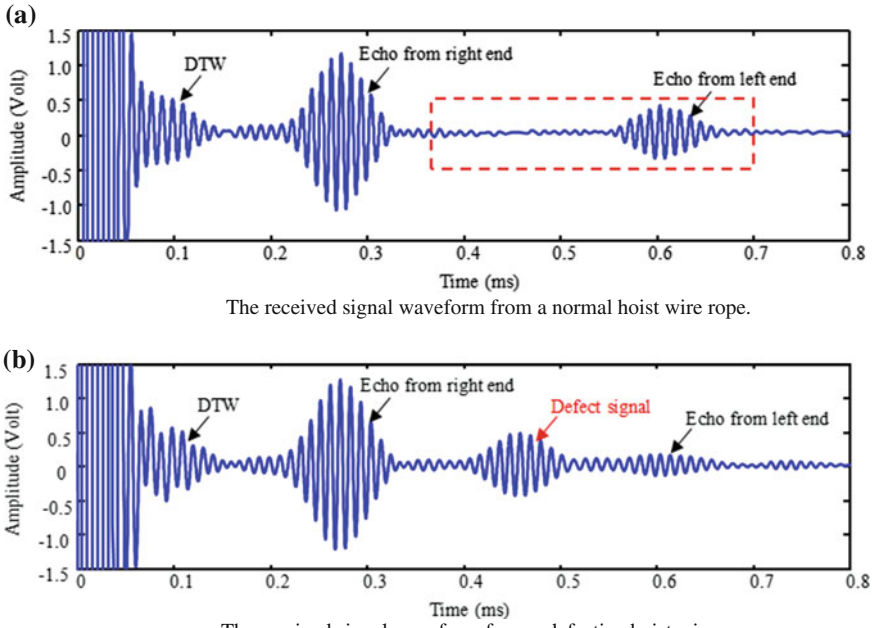


Fig. 11 The typical received signal waveforms in the tested wire ropes. **a** The received signal waveform from a normal hoist wire rope. **b** The received signal waveform from a defective hoist wire rope

The defect of broken wires as shown in Fig. 12 is apart from the left end of the rope for about 300 mm. The amplitude of wave packet "Echo from left end" reduced because a part of the guided wave energy was reflected by the defect before it propagated to the left end of rope. Reflected wave signals were recorded when the number of broken wires increased. The reflected signal waveforms are shown in Fig. 13. It is obvious that no signal was reflected at the marked position of defect in a healthy rope. When four wires were cut off in the defect location, an obvious

Fig. 12 The broken wire defects in the hoist wire rope

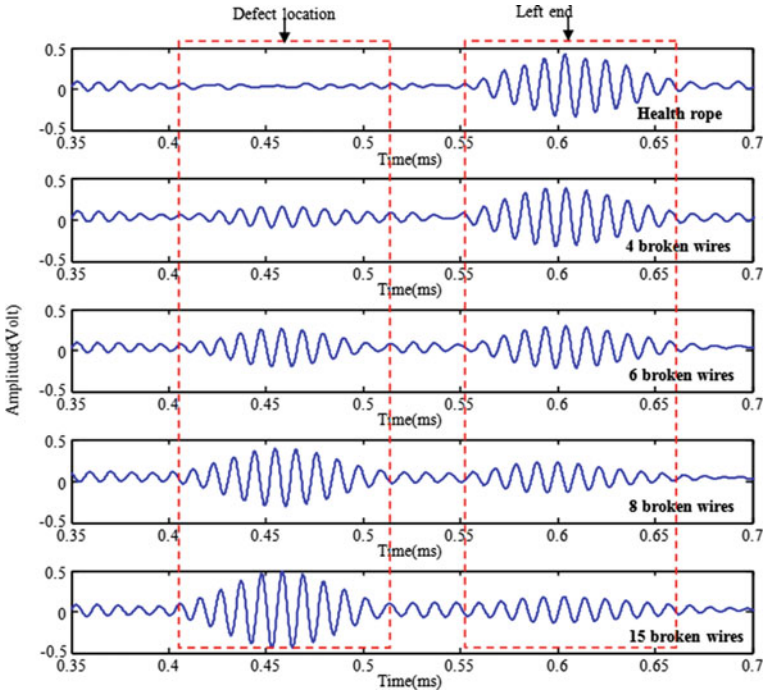


Fig. 13 The waveforms of reflected signal obtained from a healthy rope and several defective ropes with different number of broken wires

reflected wave packet was observed in Fig. 13. That mean, if a defect in the rope has more than four broken wires, which is around 2 % of the total number of wires used to form the rope, the MsS can be detected form analysing the reflected wave signal from such broken wires.

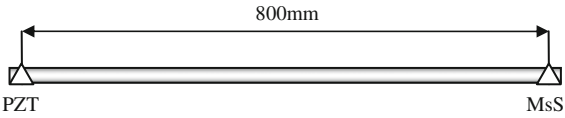


Fig. 14 The installation locations of the PZT and MsS in the rope

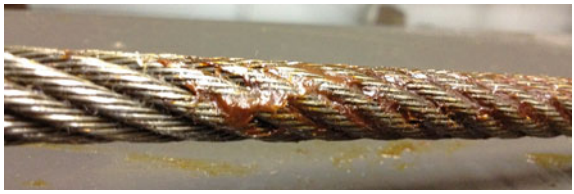


Fig. 15 The lubricating grease coated on the surface of the tested wire rope

3.2 The Inspection of Steel Wire Rope that Was Covered by Lubricating Grease

Because of improper maintenance, the elevator ropes are inevitable become dirty and makes the lubrication grease become dirty too. A proper inspection may not be possible unless the dirt or excess lubricant has been removed. Form our experiments, we found that the dirt or dirty grease will increase the attenuation of the guided wave, hence, reduce the propagating distance of the guided wave. In this study, lubricating grease was applied to the inspected rope to observe the effect on the reflected guided wave. The installation locations of PZT and MsS are shown in Fig. 14. The distance of two transducers was 800 mm apart and the pulse-catch mode was used. The surface of the wire rope was coated with lubricating grease as shown in Fig. 15.

Figure 16 depicts the received signals waveforms from the wire rope with no grease (blue color curve) and the wire rope with lubricating grease (red color curve). The attenuation of the reflected signal waveforms received from a non-coated rope as compared to the rope covered by lubricating grease was about 23 %. Hence, the range of inspection distance will be decreased due to the attenuation of the reflected signal energy. Nonetheless, the reflected signal caused by the cut end can still be clearly observed even the rope was covered by grease.

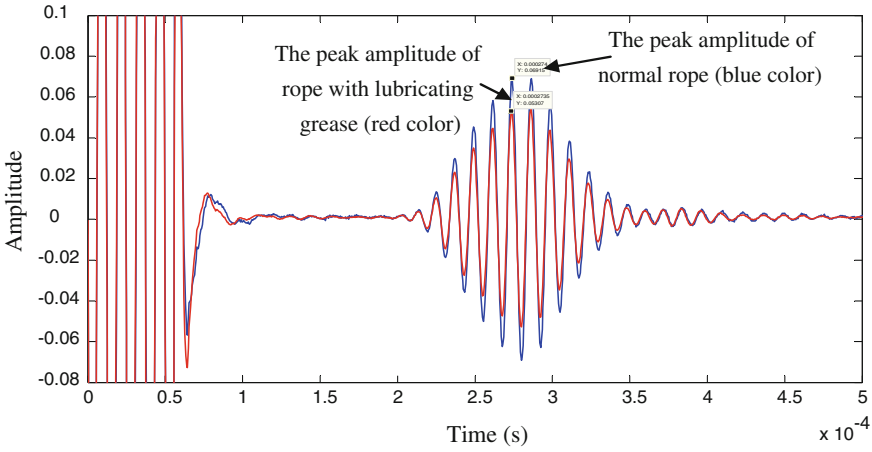


Fig. 16 The comparison on the received signal waveform from rope with/without lubricating grease

4 Conclusion

The results presented here demonstrated the effectiveness of the proposed guided wave technique for inspecting complicated structures, like elevator ropes formed by many twisted wires. The sensors made from PZT and MsS were used to emit and receive guided waves for inspecting the rope defects. The experimental results prove that the PZT sensor is more suitable for working as a transmitter/emitter, whilst, the MsS is better working as a guided wave receiver. Wire rope specimens with different number of broken wires and ropes covered by lubricating grease were tested. The findings were than used to compare with that collected from health ropes and non-greased rope. The results proved that the technique is able to detect the locations of broken wires and then determine the number of broken wires effectively. Basically the technique can detect ropes that have at least four broken wires, which is around 2 % of all wires for forming the rope. Moreover, the range of inspection distance will be decreased when applied to ropes covered by grease due to the attenuation of the reflected signal energy. Nonetheless, the reflected signal caused by the cut end can still be clearly observed even the rope was covered by grease. To increase elevator safety and reliability, it is necessary for the related industries to seriously consider the use of this new inspecting technique for steel wire ropes/cables that are hoisting heavy objects. In future, to ease the process of mounting sensors on ropes, the laser-based guide wave inspection system will be studied as proposed in published research results [5–7]. Such system can truly provide a non-contact type of guided wave emission and measurement method and an easy process in installation.

Acknowledgments This chapter was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 122011) and a grant from City University of Hong Kong (Project No. 7008187).

References

1. Kitagawa M, Suzuki S, Okuda M (2001) Assessment of cable maintenance technologies for Honshu-Shikoku Bridges. J Bridge Eng 6(6):418–424
2. Liu Z, Zhao J, Wu B, Zhang Y, He C (2010) Configuration optimization of magnetostrictive transducers for longitudinal guided wave inspection in seven-wire steel strands. NDT&E Int 43 (6):484–492
3. Wang X, Tse PW, Dordjevich A (2012) Evaluation of pipeline defect's characteristic axial length via model-based parameter estimation in ultrasonic guided wave-based inspection. Meas Sci Technol 22(2), art. no. 025701
4. Tse PW, Liu XC, Liu ZH, Wu B, He CF, Wang XJ (2011) An innovative design for using flexible printed coils for magnetostrictive-based longitudinal guided wave sensors in steel strand inspection. Smart Mater Struct 20(5)
5. Lee TH et al (2008) Single-mode guided wave technique using ring-arrayed laser beam for thin-tube inspection. NDT&E Int 41:632–637
6. Lee TH, Choi IH, Jhang KY (2008) Single-mode guided wave technique using ring-arrayed laser beam for thin-tube inspection. NDT&E Int 41(8):632–637
7. Lee J-H, Lee S-J (2009) Application of laser-generated guided wave for evaluation of corrosion in carbon steel pipe. NDT&E Int 42(3):222–227

Classifying Data Quality Problems in Asset Management

Philip Woodall, Jing Gao, Ajith Parlikad and Andy Koronios

Abstract Making sound asset management decisions, such as whether to replace or maintain an ageing underground water pipe, are critical to ensure that organisations maximise the performance of their assets. These decisions are only as good as the data that supports them, and hence many asset management organisations are in desperate need to improve the quality of their data. This chapter reviews the key academic research on data quality (DQ) and Information Quality (IQ) (used interchangeably in this chapter) in asset management, combines this with the current DQ problems faced by asset management organisations in various business sectors, and presents a classification of the most important DQ problems that need to be tackled by asset management organisations. In this research, eleven semi-structured interviews were carried out with asset management professionals in a range of business sectors in the UK. The problems described in the academic literature were cross checked against the problems found in industry. In order to support asset management professionals in solving these problems, we categorised them into seven different DQ dimensions, used in the academic literature, so that it is clear how these problems fit within the standard frameworks for assessing and improving data quality. Asset management professionals can therefore now use these frameworks to underpin their DQ improvement initiatives while focussing on the most critical DQ problems.

Keywords Information quality · Data quality

P. Woodall (✉) · A. Parlikad
University of Cambridge, Cambridge, UK
e-mail: pw325@cam.ac.uk

A. Parlikad
e-mail: ajith.parlikad@eng.cam.ac.uk

J. Gao · A. Koronios
University of South Australia, Adelaide, SA, Australia
e-mail: jing.gao@unisa.edu.au

A. Koronios
e-mail: andy.koronios@unisa.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_29

1 Introduction

The capital invested in an organisation's assets requires that maximum benefit is extracted from the assets throughout their lifecycle, which means that making sound decisions about managing the assets is critical [1, 16]. Examples of these decisions include "when should I replace this asset?" or "when should I perform a maintenance intervention?" Now, as decisions made are only as good as the information available at hand to make those decisions, the information needs to be of the required level of quality. Basing decisions on poor quality information can potentially result in great economic losses [3]. Maintaining and providing good quality information is a difficult task, and many leading asset management organisations are keen to identify areas where information quality can be improved.

To meet this need, this chapter presents seven critical information quality (IQ) dimensions (accessibility, consistency, interpretability, timeliness, accuracy, relevance, and believability) that asset managers should focus their IQ assessment and improvement programmes on. These IQ dimensions were identified through a combination of literature review and discussions with UK-based asset management practitioners from a range of industry sectors.

2 Background

2.1 Asset Management

As part of the coordinated activities to optimally manage assets, organisations must make decisions which affect the state of their assets for each of the lifecycle stages while recognising that these decisions are not independent; for example, decisions to acquire new assets are often influenced by asset retirement decisions—hence the asset *lifecycle*. Coordinating these decisions and understanding the impact of one decision outcome on subsequent decisions is vital to efficient asset management. Effective decision-making can be achieved through monitoring and capturing of information regarding key events and factors/constraints that impact on asset performance, and consequently, organisational performance. However, more data does not necessarily mean better information or more effective decisions. Providing asset managers with good quality information is, therefore, of uttermost importance. Therefore, with the wider aim of improving the way the decisions are made for all stages of the lifecycle, this chapter addresses the IQ aspects of the information which supports these decisions.

2.2 Information Quality

IQ is a multi-dimensional concept [20], and there is no general agreement on a standard set of dimensions [3]. Furthermore, drilling deeper into each dimension reveals that they do not have a commonly accepted definition [20]. The basic set of dimensions used by most authors include: accuracy, completeness, consistency, and timeliness. In order to determine the key IQ dimensions related to AM decisions, a literature review of IQ dimensions was conducted to obtain a general set of dimensions, and, importantly, their definitions (as presented Table 2). Using these key dimensions it is possible to leverage existing research on IQ assessment (see for example, [1, 16, 15]; [10]; [8] to develop tools to assess the quality of AM related information and data.

2.3 Information Quality in Asset Management

Currently, there is only limited asset management research covering IQ. This section discusses the key research in this area which is divided into two broad areas: IQ for the whole asset lifecycle and IQ for specific asset management problems. Table 1 presents a summary of this research including the source, focus on AM, and a brief description of the research.

Lin et al. [1] and [12] have developed a preliminary framework for data quality assessment related to whole-lifecycle asset management decisions. This framework has been tested using case studies with two Australian engineering organisations, and is designed to help organisations and practitioners understand AM data quality problems, identify causes, and develop solutions in accordance with three perspectives: technical, organisational and personal. The technical perspective covers the hierarchical structures or networks of interrelationships between individuals, groups, organizations and systems; the organisational perspective concerns an organisation's performance in terms of effectiveness and efficiencies; and the personal perspective focuses on individuals with issues such as job security [1]. Using this framework as a basis, and using a questionnaire-based approach, the authors identified IQ related problems in Australian asset management organisations. These problems have been combined with the results of this research in order to present a more comprehensive set of AM IQ problems (see Sect. 4).

Human factors, which form part of the personal perspective, have also been researched as part of the whole asset lifecycle. In this area, improving attitudes to IQ, improving group support of IQ, and structural solutions for improving IQ are three areas which should be addressed to improve manual data acquisition [15]. For personal attitudes to IQ, it is essential to reinforce the idea that IQ is highly valued within the organisation, improving group support focuses on ensuring that the attitudes of people are widely known and shared within their group, and structural

Table 1 A summary of related asset management information quality research

Source	AM focus	Description of research
Lin et al. [1]	Whole asset lifecycle	Presents the results of a large data quality survey related to how Australian organisations address data quality issues, and proposes a framework for data quality in asset management
Koronios et al. [12]	Whole asset lifecycle	Presents the results of 30 interviews in 2 large Australian utility companies which aimed to explore DQ issues associated with the implementation of Enterprise Asset Management (EAM) systems
Hodkiewicz et al. [9]	Maintenance performance metrics	Proposes a framework for assessing the quality of asset maintenance performance metrics, such as mean time to failure (MTTF)
Bardaki and Pramatari [1]	Shelf replenishment process decisions	Investigates the impact that the adoption of RFID has on the quality of information utilized during the shelf replenishment process and consequently on the products' shelf availability
Kelepouris et al. [11]	Supply chain tracking	Proposes a way to model supply chain tracking information in order to determine its quality with regard to its ability to support business decisions
Murphy [15]	Whole asset lifecycle	Evaluates the relative importance of human factors associated with data quality. This chapter considers the impact of attitudes, perceptions and behavioural intentions on the data collection process in an engineering asset context

solutions for improving IQ should ensure that people have the desired level of perceived control over the IQ processes [15].

Other research in the area of IQ for AM focussed on improving IQ for specific areas within the realm of asset management, such as for the "operate and maintain" stage, rather than the whole lifecycle. The "operate and maintain" stage is a key part of the asset lifecycle and there is currently no accepted method of assessing the quality of maintenance related data [9]. To address this gap, a method for assessing the quality of the data which is used to form asset maintenance performance metrics, such as Mean Time To Failure (MTTF) has been proposed in [9].

Some research focuses on IQ for decisions in the acquisition stage, such as when to order replacement parts or how many replacement parts to keep in stock [1]. These decisions are affected by the supply chain, and information concerning the location of products is needed to support these decisions. New technologies, such as RFID, can be used in this scenario to provide high quality product location information [11]. To complement this approach and to ensure high quality information is available, qualitative and quantitative methods have been developed to measure IQ with regard to how it can support business decisions [11].

3 Methodology

This section describes the methodology used to elicit the IQ problems in asset management, which included a combination of semi-structured interviews and a literature review. These methods were combined with a second literature review on IQ dimensions in order to develop the classification presented in Sect. 4. The interviews and the first literature review were used to determine the AM IQ problems in the current state of the practice in UK organisations. The inclusion of the literature review was necessary to mitigate any bias arising from the interview results, such as systematically missing certain problems, and also to provide more authoritative coverage of the problems.

Eleven semi-structured interviews were carried out with AM professionals, who either work for, or have consulting experience with, companies in a range of business sectors in the UK (covering aerospace, transportation, telecommunication, utilities, retail, chemical, oil and gas, construction and energy industry). The semi-structured interviews were based on an instrument containing four topics: AM decisions, information to support AM decisions, factors which affect AM decisions, and problems with the information used to support AM decisions. It is the results of the latter topic on problems with information which are presented in this chapter.

The IQ dimensions were not mentioned in the questions to the respondents, because there is a strong possibility that the respondents would interpret the dimensions differently; that is, attribute their own (different) definition to each dimension. Moreover, the interviews confirmed that the asset management professionals were able to relate to information problems more than abstract IQ dimensions.

4 Information Quality Problems in AM

This section presents and discusses the classification of the IQ problems both in the literature and also reported by the respondents of the interviews. These problems have been classified into the seven IQ dimensions shown in Table 2. The dimensions were selected from the list of all dimensions in the literature and according to how the definition of each dimension suitably describes the information and data quality problems in AM. For example, the definition of timeliness suitably describes the problems with out-of-date data in the AM context. It is important to note that before considering any IQ dimensions applicable to a piece of information, the information must first exist. This is often represented by the "completeness" dimension. Although described as accessibility, this point is also highlighted in [1]: "If information is inaccessible, all other qualities of it are irrelevant." Therefore, the IQ problems presented in this section, and listed in, are all related to information or data which, somewhere, actually exists.

The frequencies with which the respondents referred to different IQ-related problems are shown in Fig. 1. It is important to note that the respondents were not

Table 2 The selected definitions for the IQ dimensions

Dimension	Source(s)	Definition
Accessibility	Pipino et al. [1]	The extent to which data is available or easily and quickly retrievable
Consistency	Ballou and Pazer [1] and Wang et al. [21]	The representation of the data value is the same in all cases
Interpretability	Kahn et al. [10]	The extent to which information is in appropriate languages, symbols and units, and the definitions are clear
Timeliness	Pipino et al. [1]	The extent to which data is sufficiently up-to-date for the task at hand
Accuracy	Ballou and Pazer [1] and Wang et al. [21]	The recorded value is in conformity with the actual value
Relevance	Pipino et al. [1]	The extent to which the data is appropriate for the task at hand
Believability	Pipino et al. [1]	The extent to which data is regarded as true and credible

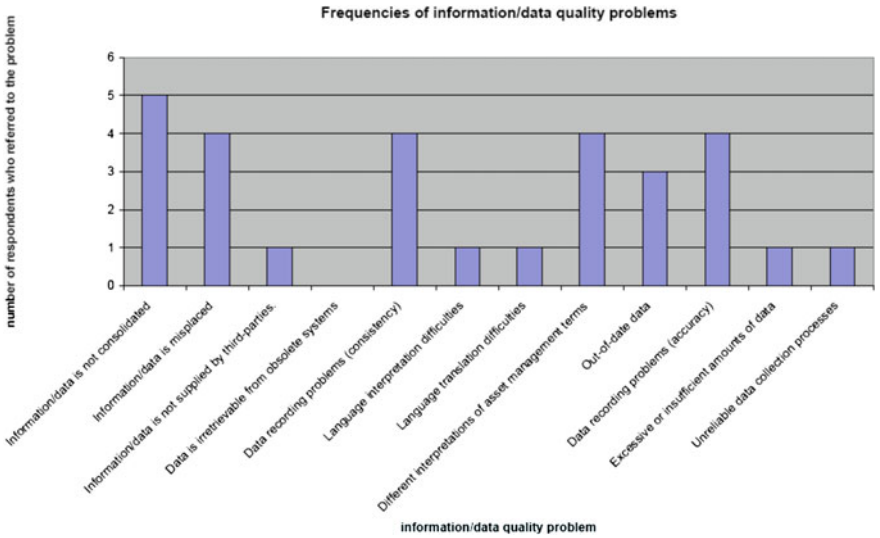


Fig. 1 Frequencies of information quality problems

given a list of problems and then asked whether they had experienced these problems. Independently and without prompting the respondents referred to the different IQ problems shown in Fig. 1. A limitation, and guard against generalising further from these results, is that respondents may have mentioned the other problems if they were prompted. This figure is therefore only an initial guide of the spread of problems throughout the different business sectors.

Figure 1 does, however, indicate that there is not a vastly different set of IQ problems for asset management companies in a range of different business sectors; a sizable proportion of the problems were experienced by more than one asset management professional. This is indicated by the six problems which were mentioned by 3–5 professionals in Fig. 1. From these results, however, there is no overarching consensus that a particular problem is dominant. At most, the problem of information not being consolidated was referred to by five respondents. The problem of irretrievable data was not reported by the respondents, and was obtained from the literature review.

We categorised these problems into seven different IQ dimensions: accessibility, consistency, interpretability, timeliness, accuracy, appropriate amount of information, and believability. Each of these dimensions is discussed in the following subsections including the associated asset management IQ problems which relate to the dimension.

4.1 Accessibility

Accessibility problems are defined as the extent to which information is available or easily and quickly retrievable (see Table 2), and there are four problems which result in information and data being inaccessible to the asset management decision-maker. These problems are discussed in the following subsections and relate to both information and data.

4.1.1 Information Is Not Consolidated

The most prevalent IQ problem arising from the interviews is that key information required for AM decisions is dispersed between multiple sources, and it is therefore difficult for the decision-maker to obtain (see row 1). Five out of eleven respondents mentioned this problem (see Fig. 1).

Information could be retained in the memory of staff within the organisation, and in this case, it is often the poor transfer of this information between people that exasperates the problem of information consolidation. In other scenarios, information also resides in multiple hardware/software systems [18] including, for example, vibration monitoring systems and Enterprise Asset Management (EAM) systems [1]. One of the main problems is that some of these systems need to provide information for decision-making as well as their primary purpose. For example, equipment maintenance and reliability data is typically needed by (i) reliability engineers for the determination of long-term maintenance strategies, and (ii) maintenance engineers and maintenance supervisors for addressing day-to-day maintenance issues [9]. These systems often do not support the first channel (i) making it difficult to extract and integrate this information for higher-level decisions. Some of these problems are as a result of the lack of an overall data architecture and management strategy [1].

Furthermore, Third-party systems are difficult and/or expensive to "open up" to extract information and share it with other applications. The end result is that decision-makers are faced with having multiple disparate sources of information and they need to spend time to collate this information or find alternative ways of making the decision.

Existing approaches to address this problem include data warehouses which hold all of the required data, and the various asset management systems must be able to provide the required data to the data warehouse. However, this approach leads to another IQ problem: an excessive amount of information. Other approaches require the synchronisation of data between systems; however, this is noted as being difficult to manage correctly [1].

4.1.2 Information Is Misplaced

Another problem which results in data being inaccessible to the decision-maker is when it is misplaced; four of the respondents mentioned this problem (see Fig. 1). Some assets operate over long periods of time having a life of many years. Information such as technical drawings is easily lost if the organisation does not have appropriate data models [1] or, generally, no effective procedure for storing the information securely.

4.1.3 Information Is Not Supplied by Third Parties

If asset maintenance tasks are outsourced by the asset operator to external companies, the required information is not always delivered back to the asset operator in a reliable process or coherent form. Similarly, with the asset acquisition process, all the information about how to own and operate the asset should be supplied to the user organisation, and occasionally, the supplied information does not relate to the acquired asset [1]. Furthermore, even if the relevant information is supplied from a third party, it could be delivered in a form which is not suitable for the user organisation [12], [1]. For example, the user organisation may require an electronic copy of the data, rather than paper-based, to enter directly into an Enterprise Asset Management (EAM) system.

4.1.4 Irretrievable Data from Obsolete Systems

For condition monitoring equipment, a lack of standards and protocols between vendors has resulted in reduced compatibility between the hardware and software of these types of systems [12], [1]. Therefore, these systems are more likely to become obsolete and be incompatible with newer versions of hardware and software.

Consequently, the data within these systems becomes very difficult to retrieve and therefore inaccessible to the decision-maker. No respondents mentioned this problem, although it is noted in the literature.

4.2 Consistency

Inconsistent data recording problems have been classified as a consistency problem, which is defined as the extent to which the representation of the data value is the same in all cases (see Table 2).

4.2.1 Data Recording Problems

When data/information is recorded inconsistently, it is very difficult for decision-makers to search and find the data/information they need and also to understand any aggregations of the data/information. The recording of labour hours is a common example where data is recorded inconsistently. For example, a machine maintainer may record labour hours on 1 day as "0.5" days, and on another day as "6 h". These may be semantically equivalent (i.e. both are defined as half of a day), but they are not syntactically consistent. Hence, aggregating and searching these values is difficult. For example, adding the labour hours could easily result in "6.5 h" rather than the correct value of "12 h", and a search for all maintenance actions taking longer than "5 h" would not always find the correct entries. An interesting side note is that this type of problem is exasperated by information systems which allow values to be entered as free text, rather than requiring structured input. A cause which is attributable to this problem is the lack of staff training, understanding and appreciation of how the data/information will be used and the consequences of poor quality data/information [1, 15]. (Note that if after 7 h of actual work, the machine maintainer records the labour hours as "0.5 days", then this is not a consistency problem. This type of problem is classified as an accuracy problem.)

4.3 Interpretability

Interpretability is the extent to which information is in appropriate languages, symbols and units, and the definitions are clear (see Table 2). Another problem, mentioned by four of the respondents, is that there are different possible definitions for terms in AM, which leads to different interpretations of the terms.

4.3.1 Language Translation and Interpretation Difficulties

Technical manuals for assets (for example, automotive and aerospace technical manuals) are often written in different languages which need to be translated for decision-makers. The decision-maker may attempt to interpret information about what components of an asset need to be replaced, and because of a mistake in the interpretation, replace the wrong component of an asset. Similarly, the decision-maker may have access to information, such as a report, which has been incorrectly translated and hence make poor decisions.

4.3.2 Different Interpretations of Asset Management Terms

The inconsistent use of definitions for the terms used in asset information is another common IQ problem within the asset management context, and was mentioned by four of the respondents (see Fig. 1). There is a wide variety of terms used in asset management, such as maintenance, lead time, takt time, refurbishment, and brake system, all of which can have different definitions and therefore can be interpreted differently. For certain terms, the problem is evident when values are calculated and recorded for these terms using one definition, and are interpreted by the decision-maker using another definition. For example, the lead-time to obtain a replacement part for an asset could be calculated using the production and shipment times, but the decision-maker could interpret the lead time as being only the shipment time. Clearly, this inconsistency could lead to poor asset management decisions. With this information, there is no data quality problem because there is no inconsistency between the values of the data, but because the interpretations of these values by different people are inconsistent, there is an IQ problem. One respondent mentioned that his organisation manages a company policy regarding the definition of terms to ensure that people use consistent definitions; however, he also noted that this is a very difficult problem to address for the entire organisation.

4.4 Timeliness

Timeliness is defined as the extent to which information is sufficiently up-to-date for the task at hand (see Table 2). A common scenario for causing information to be out-of-date for asset management decision-makers is when people develop and use their own software (mostly spreadsheets) [3] for calculating quantities, such as various asset costs. While this information may be accurate (as reported by one respondent), it is not continuously updated in the main Enterprise Resource Planning (ERP) or Enterprise Asset Management (EAM) system used by the decision-makers. For example, the information may only be entered into EAM/ERP systems each month and the entries are back-dated. Therefore, decisions which are based on information in the EAM/ERP system may be based on information which is up to

1 month out-of-date. This problem can be viewed from the two perspectives of technology and people, because information systems which are not easy to integrate exasperate the problem, and staff may have no incentive to keep the information in other systems up-to-date. Furthermore, this is both an information and data quality problem because data or information may be stored in the individual systems before being updated in the main ERP system.

4.5 Accuracy

Accuracy describes the state when the recorded value is in conformity with the actual value (see Table 2). Four respondents reported data recording problems which result in a discrepancy between the actual "real" value and the recorded value (see Fig. 1). One cause of a data recording problem is when the person recording the information makes a mistake and records the wrong value (this is a DQ problem); another is when a person knowingly enters false information. This could range from entering an incorrect temperature value (data) or entering false information about the cause of failure of an asset. These problems, therefore, occur with both information and data.

Usually, people knowingly record information which may be incorrect, when they are required to record the information and the actual (correct) value or truth, for the information, is not known. A common example of this is when information systems require the cause of failure of an asset to be recorded, but the user does not know the exact cause of failure. Validation checks can prevent the user from continuing to enter other useful information and so the user enters any value which passes the validation check regardless of whether it is the correct value. Similar to the inconsistent data recording problems, one of the causes of this problem is the lack of staff training, understanding and appreciation of how the data/information will be used and the consequences of poor quality data/information [1, 15]. This problem could also be attributed to the information system related to how it enforces validation checks.

4.6 Appropriate Amount of Data/Information

If there is an excessive amount of data, then the decision-maker will find it difficult to locate the data needed to make certain decisions. In the asset management context, excessive amounts of data can easily be generated and one example of this situation is in asset fault reporting. One respondent mentioned that, in information systems, faults are often recorded using fault codes which link to a description of the fault. However, when asset faults are recorded, it is not always easy to find an existing code which exactly matches the fault, and, therefore, the fault is recorded using a new fault code. This results in a large number of fault codes with no general

set from which the decision-maker can base sound decisions. The problem of appropriate amounts of data/information can occur with data or information; although, it is more common for this problem to be data related, especially if data is captured from asset condition sensors.

Another contributor to the problem of inappropriate amounts of data, which could either be an excessive or insufficient amount [9], is the lack of "why-knowledge" of data collectors. This is the extent to which the data collectors understand why the data is being collected and how it will be used, without such knowledge it is likely that incorrect amounts of data will be collected [9].

4.7 Believability

Asset condition information is not always believable, which is the extent to which data/information is regarded as true and credible. Direct rating or distress surveys are among the existing approaches for the inspection and evaluation of the current physical condition of assets, however, the results of these are based on a subjective visual inspection process [7]. Without experienced inspectors who follow the correct procedures to judge the condition of an asset, the resulting asset condition reports are not trustworthy, and hence the decision-maker does not always believe and have confidence in the information. This problem was reported both in the literature and by one of the respondents.

5 Conclusions and Further Research

The aim of this research was to identify and classify the DQ/IQ problems which affect AM decisions. Twelve problems have been identified, which are categorised into seven different IQ dimensions. Further research can therefore combine this classification with the existing research on information and data quality to develop suitable techniques for assessing DQ/IQ within an AM context. Moreover, these dimensions can be used to form the basis of assessment and subsequent data/IQ improvement practices within AM organisations.

An interesting aspect is that a sizable proportion of the problems are experienced by multiple organisations in different business sectors; however, five of the problems were referred to by one asset management professional. Only one problem was noted in the literature which was not mentioned by any of the respondents. This shows that the literature is focussing on the relevant areas, and also that UK companies are experiencing similar data/IQ problems as Australian organisations, because a significant proportion of the problems reported in the literature were obtained from a survey of Australian AM organisations [1].

Using these dimensions it is possible to leverage existing research on IQ assessment (see for example [1, 16, 15]; [10]; [8] to develop tools to assess the quality of AM related information and data.

Using this work as a basis, we aim to develop an IQ audit tool for use with AM organisations to assess their current level of information and data quality. The current levels can be compared to the ideal levels in order to identify areas where the information and data quality needs to be improved within the organisation. Further research is therefore required to determine suitable ways to measure the information and data quality dimensions critical to asset management organisations and determine the ideal level of IQ.

References

1. Ballou DP, Pazer HL (1985) Modeling data and process quality in multi-input, multi-output information systems. Manag Sci 31(2):150–162
2. Bardaki C, Pramatari K (2007) Assessing information quality in a RFID-integrated shelf replenishment decision support system for the retail industry. In: 12th international conference on information quality, p 302
3. Batini C et al (2009) Methodologies for data quality assessment and improvement. ACM Comput Surv 41(3). http://www.scopus.com/inward/record.url?eid=2-s2.0-70349690001&partnerID=40. Accessed December 17, 2009
4. Bovee M, Srivastava RP, Mak B (2003) A conceptual framework and belief-function approach to assessing overall information quality. Int J Intell Syst 18(1):51–74
5. Gao J, Baskarada S, Koronios A (2006) Agile maturity model approach to assessing and enhancing the quality of asset information in engineering asset management information systems. In Proceedings of the 9th international conference on business information systems (BIS 2006), pp 486–500
6. Gao J, Lin S, Koronios A (2006) Data quality in engineering asset management organisations—current picture in Australia. In: The 2006 international conference on information quality, p 18
7. Hegazy T et al (2008) Ultra mobile computer system for accurate and speedy inspection of buildings. Annual conference - canadian society for civil engineering, In proceedings, pp 456–464
8. Hill G, Price R, Shanks G (2004) A semiotic information quality framework: applications and experiments. proceedings of the IFIP international conference on decision support systems (DSS 2004): Decision Support in an Uncertain and Complex World. Prato, Italy, pp 658–672
9. Hodkiewicz M et al (2006) A framework to assess data quality for reliability variables. Gold Coast, Austrialia, In Proceedings of the world congress on engineering asset management (WCEAM)
10. Kahn BK, Strong DM, Wang RY (2002) Information quality benchmarks: product and service performance. Commun ACM 45(4):184–192
11. Kelepouris T, McFarlane D, Parlikad A (2007) Developing a model for quantifying the quality and value of tracking information on supply chain decisions. In proceedings of the 12th international conference on information quality (ICIQ-07), Boston
12. Koronios A, Lin S, Gao J (2005) A data quality model for asset management in engineering organisations. In proceedings of the 10th international conference on information quality (ICIQ 2005) pp 4–6.
13. Lima LF, Macada AC, Koufteros X (2007) A model for information quality in the banking industry—the case of the public banks in Brazil. In: The 12th international conference on information quality

14. Lin S et al (2007) Developing a data quality framework for asset management in engineering organisations. Int J Inf Qual 1(1):100–126
15. Murphy GD (2009) Improving the quality of manually acquired data: applying the theory of planned behaviour to data quality. Reliab Eng Syst Safety 94(12):1881–1886
16. Ouertani MZ, Parlikad AK, McFarlane D (2008) Asset information management: research challenges. In Proceedings of the 2nd international conference on research challenges in information science, RCIS 2008, pp 361–370. http://www.scopus.com/inward/record.url?eid=2-s2.0-57349094377&partnerID=40
17. Pipino LL, Lee YW, Wang RY (2002) Data quality assessment. Commun ACM 45 (4):211–218
18. Soma R, Bakshi A, Prasanna V (2007) An architecture of a workflow system for Integrated Asset Management in the smart oil field domain. In Proceedings of the 2007 IEEE congress on services, services 2007. pp 191–198. http://www.scopus.com/inward/record.url?eid=2-s2.0-46849085383&partnerID=40
19. Stvilia B et al (2007) A framework for information quality assessment. J Am Soc Inform Sci Technol 58(12):1720–1733
20. Wand Y, Wang RY (1996) Anchoring data quality dimensions in ontological foundations. Commun ACM 39(11) pp:86–95
21. Wang RY, Storey VC, Firth CP (1995) A framework for analysis of data quality research. IEEE Trans Knowl Data Eng 7(4):623–640

Asset Data Quality—A Case Study on Mobile Mining Assets

M. Ho, M.R. Hodkiewicz, C.F. Pun, J. Petchey and Z. Li

Abstract Good asset management decisions involve balancing cost, risk and performance requirements. Raw data on maintenance costs (a major contributor to total costs) and for estimating risks associated with asset failure is stored in an organisation's Enterprise Resource Planning (ERP) system. However as this chapter demonstrates asset data is often erroneous, lacking requisite detail and therefore not fit for decision support. This chapter describes a project to clean data stored in computerised maintenance management systems (CMMS) that form part of ERPs. It looks in detail at the cleaning process, identifying key issues and developing of a set of recommendations for improvement. The major issues identified are to do with poor practice in assigning work to appropriate subsystems and maintainable items, ineffective use of standard text to describe work, inconsistent use of codes describing the type of work, and inability to identify suspensions and actual asset usage hours from the stored data. While focussing on asset data from mobile mining assets, the problems identified are similar in other sectors. Despite these issues, much of the required information is available once the data has been cleaned and forms a resource for the mining industry to assess how asset reliability and costs are changing with the introduction of new developments such as autonomous mobile equipment.

Keywords Asset management · Data quality · Computerised maintenance management system · Data cleansing · Mining · Heavy mobile equipment

1 Background

The mining industry is asset-intensive, requiring substantial upfront investment in physical assets, particularly heavy mobile equipment, to access the ore body and subsequently move material (ore and waste) direct to market or onto further

M. Ho · M.R. Hodkiewicz (✉) · C.F. Pun · J. Petchey · Z. Li
University of Western Australia, Crawley, WA, Australia
e-mail: melinda.hodkiewicz@uwa.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_30

processing. As an indication of size, in 2010–2011 the Australian mining industry made capital expenditures of A$57.1b [1]. The assets purchased as part of this investment have to be maintained over their life to ensure return on capital. Collectively the asset base of the Australian mining industry has replacement value well in excess of this number and substantial (if unpublished) dollars are spent maintaining it annually. This maintenance work is mainly (but not exclusively) recorded in the mining companies' CMMS.

Data to support asset management decisions, particularly maintenance data, is widely acknowledged to be incomplete, erroneous, out-of date, embedded within significant amounts of meaningless data, incompatible with analysis tools and lacking in requisite detail [2]. This chapter reports on work looking at the raw data in the CMMS, specifically at a sample of 1,086 heavy mobile equipment assets. It describes the data cleaning process, identifies major issues and provides relevant illustrations. One of the aims of the chapter is to generate discussion between those in the maintenance community responsible for generating the maintenance data (data collectors), the data custodians (IT groups) and users of the data, primarily asset management professionals and reliability engineers.

This work is part of a larger program to understand the historical reliability of mobile mining assets and to identify what data might need to be stored as part of a mining industry reliability database. The existence of industry-specific databases is widespread in other sectors and has supported a steady increase in equipment reliability through design improvements. Examples include OREDA in the oil and gas sector, T-book in nuclear, SPIDR for electronic and electro-mechanical components, PERD in chemical and FERMS for aerospace.

This project started at a time when a number of mining companies were implementing new Enterprise Resource Planning systems with associated changes to their CMMS. In parallel there has been substantial investment in asset management teams, asset management education, the development of maintenance tactics, improved maintenance planning and operational practices and the introduction of an asset management ISO standard. The work reported here draws on data collected prior to these improvements and so should not be regarded as where the industry is now, but a useful comparison to where it has been in the decade leading up to 2011 based on a sample of 1.34 million work orders with associated costs of $3,614 million.

There has been considerable work in the IT fraternity on data quality assessment and numerous books and articles on the theory [3–11]. However these publications deal almost exclusively with issues in relational databases and with what is described as structured and semi-structured data. The useful data for this project, useful in the sense that it is what will provide input to reliability assessments, is stored as unstructured data. Comparatively little work has been published on protocols for cleaning unstructured data. In this study 67 % of work orders based on a sub-sample of 319,339 work orders on 367 individual mobile assets from three different classes of equipment (trucks, loaders and excavators) had unique work order descriptions, i.e. they are free text and do not have a standard field describing the task that needs to be done or was undertaken. The remaining 33 % use a standard text that is repeated

at least twice. This free-text field contains information on what work was done and to what subsystem. Deciphering, cleaning and coding the unstructured text in the work orders is the main challenge discussed in this chapter.

2 Method

There are few examples in the asset management literature on data quality and these tend to focus on frameworks and the need for better data quality rather than on the methods by which it can be achieved [12–15]. There is an emerging literature looking at improving how asset data is collected [16, 17] but there are no major contributions detailing how existing asset data can be cleaned efficiently and what the key issues are.

The following steps set out the method used to clean the data in the hope of improving transparency around the processes used and ensure the work can be replicated and improved. For the purposes of this analysis five pieces of information are required from a work order to identify factors influencing the reliability of major systems on mobile mining assets.

- The date an action was taken, particularly actions relating to the repair or removal of a subsystem or maintainable item.
- The specific mobile asset (e.g. truck, dozer) and associated sub-system or maintainable item on which the work was done(e.g. engine replacement, radiator, truck tray).
- A clear description of the action or work done.
- The asset's average usage hours.
- If the equipment was removed from service while still functional.

Other useful information includes indication of whether the work: is done in response to a breakdown, is a maintenance tactic, results from a maintenance tactic, or none of the above. An indication as to whether the work done was planned or unplanned. The cost associated with the work order.

Work order data was collected on mobile assets from a range of organisations. Individual data sets are received as an Excel file containing fields downloaded from the CMMS. The process used to clean the data uses the following steps.

1. The raw data files are separated by site and asset class (truck, excavator, loader etc.).
2. A data cleansing rule file is developed. This contains conditions and actions to be performed if the conditions are met. Conditions are presented in the form of keywords and logic statements (e.g. contains, equals, does not contain, does not equal) while actions performed write, delete or copy data into specified locations.
3. The rule file is used to identify (a) systems of interest e.g. engines and (b) code actions (repair/replace/inspect). An extract of a rule file to identify fields

relating to engine replacement contains the following code (where d represents the column containing short text). A typical rule set will have 70–125 rules. *[d-has-eng/d-excludes-enge/d-excludes-enga] [d-has-o/haul/] [d-has-overhaul/] [d-has-o/h/d-excludes-o/he/].*

4. The rule file is executed using a customised MATLAB script on the raw data producing an initial pass of clean data. This program was developed as part of a purpose built data cleansing tool for this project and rule files, once developed, can be reused on similar assets.
5. Failure and suspensions are initially established based on work order type codes but may be corrected if necessary via visual data examination.
6. Data about the start/end dates for each unit and Maintenance and Repair Contracts (MARC) contract/labour agreement data are added, these come from separate data sources.
7. Work orders are sorted by asset, sub-system and date producing a chronology.
8. Assets which are maintained by MARC contracts over most of lifetime are removed from the data set since limited data is stored in the CMMS about activities conducted under MARC contracts.
9. There is then a manual check of rejected data (for records which should not have been rejected) and "good" data (to reject minor fixes and ongoing symptoms of failures, irrelevant data and fix work order types and work order codes (called PM codes), as necessary) of remaining assets.
10. Calculate calendar days between failures/suspensions and between planned maintenance events. Running hours are then determined based on calendar days and average usage hours.
11. Time intervals between planned maintenance events are reviewed and work order type codes adjusted (if justifiable). Information on external suspensions (from midlives etc.) is added and hours between failures/suspensions adjusted.
12. The following records are removed.

    a. Machines with no suspensions if its usage duration implies there should be at least one (inference of missing data) from the data set.
    b. Work orders associated with a symptom (e.g. transmission overheating) if a cause/action follows closely (e.g. replace transmission). This avoids repetitive entries for the same failure.
    c. Work order associated with the same ongoing symptoms (pending a review of costs) if less than 3 weeks have elapsed between the symptom work orders. An assumption is made that previous work orders pertaining to these symptoms relate to monitoring or minor corrective activities not work associated with a failure.
    d. Minor items not within the boundary scope of the subsystem under analysis (although these should have already been captured in the rule file), e.g. hoses, bolts, mounts etc.
    e. Minor repairs based on cost (using rules such as "if less than a given percentage of total replacement cost").

f. Data points following a MARC contract as maintenance activity performed under the MARC contract is not visible as separate work orders.
g. The first data point relating to work in cases where the start date is uncertain.

13. Finally, the clean data set is reviewed and statistics developed using Excel and the statistics package R.

3 Results

This section describes the challenges in developing a data set that is fit-for-a-specific-purpose based on data from a sample data set of more than a million work orders on mobile mining equipment over the period of a decade. The purpose is to obtain distributions of life data for sub-systems and major maintainable items on heavy mobile equipment used in the mining industry. The desired data set should contain a unique record for each work item relating to planned and unplanned activities performed on a sub-system or maintainable item over its life. The record should be clearly identified at the right level in the asset hierarchy, have an associated data, cost and describe the work that was done. In the process of cleaning data to make it fit for the purpose described, a number of challenges were noted.

3.1 Asset Allocation

Asset allocation, the identification of the mobile asset to which the work is being done, is done well. Almost all work orders are associated with a specific truck, loader, grader etc. In general the asset class is part of the asset identification for example TK001, where TK is a code to represent the truck asset class. Seldom is there a separate asset class variable (column), where a class indicates a haul truck, grader, water truck etc. Many organisations have a separate variable (column) describing the asset, generally this includes the make, model and class. When searching this variable one needs a complex set of rules. For instance a haul truck may appear in records coded as *Haul Truck, H/Truck, Rear Dump Truck or Truck*. The search must also exclude the terms fuel truck, lube truck, service truck, water truck and dewatering truck. This process is further complicated by the use in this variable of additional words relating the make and manufacture of the asset. For example, on a single site there can be as many as 348 unique entries for the make and model of various trucks.

3.2 Sub-system Allocation

Work orders are assigned to an asset's functional location. However, historically the assigning of work to an appropriate sub-system (such as engine, drive train, electrical), or a maintainable item (tray, cooling system, ignition system, cab) has not been done consistently. This is illustrated in Table 1. It shows that for 73,942 records assigned to trucks only 29 % have been assigned to the appropriate sub-system and 21.5 % to the maintainable item in that sub-system. This makes it extremely difficult to find data on particular sub-systems, for example, as 71 % of the records are stored at the top level of the hierarchy.

The only way to identify records containing important sub-systems is to data mine the Short Text field which indicates what action was taken. This situation forces the analyst to rely on the Short Text in order to determine which part of the machine was worked on. Given the diversity of the Short Text entries, this can be both inefficient (in the sense that it requires a larger number of rules to clean) and inaccurate due to the fact that the Short Text may be non-specific. Challenges with deciphering the Short Text field are described below. Another impact of this is that a search based on the coding associated with identifying the sub-system or maintainable item, when it is available, is searching less than a third of the database.

For all sites, there are multiple ways of referring to the sub-system, and also the same maintainable item within the Short Text field with no apparent standard terms. For example a search for *"Driveshaft"* in the Short Text field also needs to search for variants such as *Drive shaft, main shaft, d/shaft, driveline, drive line, drvline, drv line, d/line, and drive axle*.

Some of the impacts of past practices in the use (or not) of sub-system and maintainable items codes are shown in Table 2. This compares a search on the original data set using key words of *"Torque"* and *"Converter"* in the Short Text field for the asset class Trucks. In the top example only a fraction of the work orders that should be allocated to torque converter are actually located (the cleaning process identifies more relevant records), leading potentially to grossly inaccurate estimates of sub-system life and incorrect costs.

The second example in Table 2 illustrates a different, but common, issue where a sub-system is identified but it is mainly the maintainable items associated with that sub-system that have failed, not the sub-system itself. For example, work orders containing engine in the Short Text may pertain to covers, guards, mounts, gauges etc. rather than the engine itself. In order to focus on the reliability of the engine,

Table 1 Estimation of number of maintenance records assigned to a sub-system (the remainder are allocated to the asset)

Asset class	Number of records	% of records assigned to system (trucks)	% of records assigned to a sub-system	% of records assigned to a sub-system and maintainable item
Trucks	73,942	99.7 %	29 %	21.5 %

Table 2 Examples illustrating data quality errors

	Number of records	Cost of work in record set
Records identified from an original data set relating to key words "Torque" + "Converter" in Short Text field	527	$409,681
Records on torque converters in the cleaned set	1,361	$561,649
Records identified from an original data set relating to the key word "Engine" in Short Text field	6,256	$33.03 m
Records on engine replacement, failure or mid-life refurbishment in the cleaned set	284	$26.38 m

these other components need to be excluded in the data cleansing process otherwise the engine life is dramatically underestimated. At one site a significant percentage of the work orders on the engine related to issues with the cover and mounts, only 284 of the original 6,256 records pertain to the replacement or failure of the engines.

A very common issue for machines with multiple numbers of a same part (e.g. turbochargers) is that there is no indication of which part was changed. This is made even more complicated when multiple parts were changed out. For example Truck001 has Short Text: "*Replace blown turbo*", or Truck002 has Short Text "*changeout failed turbos*". In the second example, the position/number of turbochargers which were changed out is not known. Furthermore, some parts may be preventatively replaced along with failures under the same work order. When this happens which were failures or suspensions and the specific locations of the failures/suspensions can also be unclear.

In cleaning the data, where there were a number of the same part in a machine and uncertainty as to which was worked on, assumptions had to be made. If the work order in question occurred less than a month after a previous work order on the target parts, then a chronic failure of the previous part was assumed. Else, the part was assigned the specific location of the longest surviving part.

3.3 Work Order Description

Often times a work order describes a symptom with no indication if any action was taken, and if so, what. For example, Loader003 has the Short Text: "*transmission overheating*". In such an instance, the cost of the work order was used to infer whether it was merely an inspection or further work was done without updating the Short Text. For example, Loader9 has in the Short Text "*Inspect lower drive shaft*" but the cost is $5,320 which is roughly the cost of a new drive shaft and so the work order is unlikely to be only an inspection.

Both examples above heavily relied on cost to tell whether any work was done. If the Short Text fails to identify what was done, the cost can give an indication of

whether it was a minor repair, major repair, or a full replacement. However, in a number of cases, the cost is reported as zero, despite the Short Text indicating that a major repair was done. In other cases, the cost is too small to even cover labour costs, suggesting that very small costs are also incorrect. This can make it difficult to judge the amount of work that has been done in a particular work order. The combination of not knowing the cause or action taken and a lack of confidence in the cost field is particularly difficult where a problem could have been due to multiple causes. Overall, the number of records where analysts had very little information and it was felt to be a significant failure was small, perhaps 5 % of the total records processed. In the other 95 % there is enough information to make reasonable assumptions.

One of the main issues is with work orders which cancelled or moved planned maintenance activities without any indication on the records. As the "cost" field could not be very well trusted, even a $0 work order may not necessarily have meant a work order was not completed at that time.

The use of short hand in the Short Text can sometimes lead to confusion as to what action was taken. For example, the abbreviation "*r&r*" could refer to "*remove and repair*", "*remove and replace*" or "*remove and refit/reinstall*" (for access to other parts). The first two meanings would mean a failure/suspension whilst the last interpretation would not. Sadly, sometimes the Short Text was just complete gibberish, with no indication of what was done. For example, Track Dozer 1 Short Text: "*read text*", Truck 1 Short Text: "*fgfgfgfg*" or Loader19 "*No radiator planned maintenance interval".*

One of the more challenging aspects is in dealing with records that appear to give appropriate information but do not make sense or provide contradictory information. This is illustrated in Table 3. These types of contradictions when the cost indicates that work different than that specified in the Short Text was done are particularly difficult to identify using rules. In this study, these situations were identified manually.

Table 3 Examples of Short Text fields in which description is misleading or contradictory

Short Text	Contradictory information	Interpretation of what actually may have occured
C/O SX Engine C and D Checks-HRS	Cost is given as $96,440	Only C and D service was performed, the engine was not C/O (changed out)
PM 1,000 Hr Inspection	Cost is given as $31,637	Inspection was stated but cost implies significant unspecified work was performed
Change LH Front Final Drive on Hrs	PM code given as breakdown	Short Text states fixed interval replacement however W/O type is breakdown

3.4 Work Order Type and PM Codes

For data cleaning the Work order type and PM code, when available, is used to determine whether the work order was generated due to failure or part of a preventive maintenance scheme. In a number of cases, the Work order type and/or PM code will identify a sub-system as being preventatively replaced, yet work orders prior to the supposed preventative replacement suggest the sub-system has actually failed. This kind of scenario occurs when a problem is identified, and there is sufficient time to plan a replacement of the component that is about to fail. As such, this work order becomes part of the planned maintenance, even though failure of the part is imminent. This makes data cleaning problematic, as each point needs to be manually checked to confirm whether the Work order type and PM code is accurate or not. Due to the large number of observations this can be time consuming, and there is a possibility that some points will be missed. The example given in Table 4 illustrates the complexity of the issue. If the only point considered is the engine change out on 17/12/2003, an analysis might conclude that the engine was preventatively replaced. However, by looked at the history as well as the time since the last replacement, it can be deduced that the engine has actually failed. The coding of "planned" may be technically correct in that the work order was submitted prior to the planning window and therefore done in that week as planned work but from a reliability perspective this should be considered a failure and not a suspension given the prior work orders.

There is evidence of incorrect use of work type codes, with work coded as "preventative" that is not a result of work relating to maintenance tactics associated with fixed time/interval repair/replacements/inspections or condition based maintenance. This impacts the coding of failures as either failures of suspensions, an important consideration in life data analysis. Of particular concern are potential cases of masking breakdowns as suspensions as shown in Table 5.

Table 4 Example of assigning work from a failing engine to planned work

Date	Short Text	WO type	Cost
6/12/2003	Low eng power	Breakdown	~$150
10/12/2003	Engine oil leak at rear of engine	Breakdown	~$75
16/12/2003	Red eng monitor	Breakdown	~$300
17/12/2003	Engine change out	Planned	~$96,000

Table 5 Examples of breakdowns marked at suspensions (preventative replacements)

Short Text	W/O type	Our comment
Engine change out @ 15,000 HRS	Preventative	This work order occurred 2,992 h after previous preventative replacement a long way short of the desired 15,000 h interval
Replace leaking RH rear strut	Preventative	The leaking (failed) system is listed as a PM

3.5 Running Times and Usage Hours

The work orders do not record running hours for the mobiles machines. While these may be recorded in other systems it remains a challenge to reconcile the date and time of the work order with the life of the unit. In the absence of this information, for this project average usage hours are calculated using regular maintenance work orders (e.g. lubrication and mechanical services) with service intervals and from examining the machine maintenance reports which are in a separate file. Most sites indicate the service interval in the Short Text field however there are some that do not, for example Truck 4 *"Scheduled PM and Lube Service"*.

3.6 Missing Records

On several occasions, there were cases of some machines going for a month or two without receiving any maintenance. It is possible that the machine was not used for this period, although for a haul truck that has seen regular use before and after the gap this may be unlikely. In these cases even if there were no failures during that time there should still have been the regular lubrication service of the asset.

Often no description was given on what was done in labour and contract agreements. While the site might maintain a particular sub-system, the existence of blanket contracts and the absence of work orders for extended periods indicates that work done under contracts is not being recorded in the CMMS. This potentially extends the time to failure/suspension of the data point after resulting in an over-estimation of life. Conversely, there are occasions when there is a record of the contract but no indication of what work was done, for example, Shovel 2 "*Contract Labour for 2 PM $60,000*".

Similarly, MARC contracts did not show what items were covered under contract (and sometimes no duration). In such situations, where there was a work order directly after the labour agreement/contract with a very large time before failure/suspension (well in excess of the planned maintenance cycle usage hours for the part), it was assumed the labour agreement/contract had the target part worked on and that the following data point was deemed void.

3.7 Unrecorded Entries

Hidden entries refer to components that are changed out as part of a regular maintenance scheme, but are not recorded separately. The most common example is the engine midlife (also called engine service or engine maintenance), which may cover components such as the fuel injectors, water pump, radiator and turbo charger. For the midlife, it is common to see one entry for the midlife itself without

any separate entries for other components that are changed out as part of the midlife. This results in time consuming cross checking of midlife change outs, machine rebuilds etc. to establish suspension status.

4 Discussion

The aim of this report is to highlight some of the key issues found when cleaning data for mobile mining equipment assets. While there are issues, there is considerable information contained in these records that can be used to provide a baseline for the industry for the decade to 2012. This baseline is important for two reasons. The first is that significant investment is going into improving the collection and coding of maintenance data at a number of mining organisations. The results of these improvements can be compared with the data resulting from this study. The second is that moves to autonomous equipment will likely result in changes in failure behaviour of the assets and their associated sub-systems. Already there is anecdotal information that the more predictable operation of the autonomous units is leading to less wear and tear on the sub-systems. The life data on the sub-systems that will be produced as part of this work can be used for comparison with the autonomous sub-system life data and also failure modes.

Four main technical issues highlighted by this work are:

1. Too may work orders are assigned to the top hierarchical level (e.g. the truck, the loader, the grader) and not the appropriate sub-system or maintainable item. In many cases a hierarchical structure exists but for various reasons use of it has not been enforced.
2. Because of item (1) information about what work was done and to what asset is only available in the work order Short Text field. This field is free-text, making it almost impossible, without use of cleaning tools and processes similar to those used here, to extract data with confidence.
3. There is a significant manual element in the latter steps of cleaning and good contextual knowledge of the asset type, expected failure modes and maintenance work is required. The team involved in this cleaning all had mechanical and/or electrical engineering backgrounds as well as maintenance and reliability experience. To clean 42 sub-systems on different assets (trucks, loaders, dozers) took $\sim$1,500 h over a 3 month period.
4. From a reliability perspective, knowing if the sub-system or maintainable item has been preventatively replaced or has failed is crucial in determining statistical parameters on the sub-system or maintainable item groups. Failure to account for suspensions (preventative replacements) leads to gross errors in estimating statistical parameters such as the mean time between failures (MTBF).

Ideally, from a reliability engineer's perspective, it should be possible to search by function location for the sub-system or maintainable item that is to be the subject

of analysis. Once all the work orders for that sub-system have been selected, it should then be possible to filter to determine which entries were:

- Failures that required replacement or repair,
- Work associated with defined maintenance tactics (time/interval based replacement, or resulting from condition monitoring or inspection activity),
- Work identified by the operator or maintainer that is not covered by maintenance tactics, and
- Sub-systems or maintainable items that are functioning bur preventatively replaced (suspensions).

The asset usage hours between work orders would also be provided, allowing the data to be cleaned efficiently and the time between failures and order of the failures (1st failure, 2nd failure etc. for repairable maintainable items) to be computed.

It is worth noting that the CMMSs from which this data was obtained are not the only source of information on mobile asset cost and maintenance data. Much information is also available from the Original Equipment Manufacturers both through the MARC contracts and because they perform many of the rebuilds and major repairs. In addition, each mining company has a heavy mobile equipment engineering group which has traditionally kept track of costs and equipment performance. This has largely been done using Access databases and Excel spread sheets. The challenges with this are that this data can be highly customised to the needs of a particular engineer and it is not easily accessible to others. There has been a move amongst the major mining companies to make the ERP and associated CMMS module the repository for corporate data and to ensure there is one accurate version of the raw data. This work supports that vision. However for the purposes of the reliability study that is the focus of the larger project being undertaken here, the existence of these other sources of data allows for the possibility of triangulation and hence validating reliability data.

There are organisational reasons why maintenance data issues are so widespread across organisations and why projects like this are so difficult to do. At a simple level, part of this is to do historically with the subordinate status of maintenance particularly in the mining industry. However maintenance groups can be part of the problem. They often operate as a closed shop, or tribe as described by [18], preferring less rather than more transparency around their activities and developing a language and culture of their own. This tribal nature also makes it difficult for engineers, particularly those without maintenance experience, to gain access to and trust from maintenance groups. This access is important in understanding what is done and why.

Over the years the issues with asset data quality have become increasingly apparent, especially with the growth of programs like Six Sigma which are heavily dependent on data. For many years the industry has got by with decisions based on experience rather than on evidence, but in today's competitive global economy and focus on managing risks and optimising operations, data is a necessity. Having said this, until recently senior managers who open the lid on the asset data quality are quickly overwhelmed by the scale and complexity of the issue and the potential

costs and time involved in making material changes. No one gets recognised for improving asset data quality; in part because it has not been able to be measured. Fortunately, this is changing with massive organisational efforts to improve the way maintenance data is captured and coded, and in developing appropriate maintenance tactics, executing planned maintenance work and managing the reliability and costs of the assets. Considerable effort is also going into training mining sector asset management personnel in the type of evidence-based decision making [19] that draws on the data that is the focus of this project.

5 Recommendations

The following recommendations are made for future data collection to avoid the issues described above. As mentioned in the introduction, a number of these steps are underway in mining companies, and it is expected that the frequency of the issues reported here to decline with time and continued focus.

1. Develop standardised formats and vocabulary for each field.
2. Common terms should be used in the Functional Location field rather than many different terms for the same sub-system, maintainable item and component.
3. The input of make, model and class of mobile equipment should allow for ease of sorting.
4. Where possible, require standard entries in the Short Text field. It is possible to have a set of pre-defined terms for the majority of the work. A guide to suitable standard entries can be developed by looking at the historical data. As far as possible the entries, particularly for the Short Text field should be in the language of the operators and maintainers who generate the work notifications and reflect what is actually observed rather than the failures engineers think might happen.
5. A separate field should be created to indicate a suspension. This is when a sub-system or maintainable item is removed but still functions. This preventative replacement occurs for a variety of reasons, such as interval-based replacement tactic, a new/alternative design is being tested, or another sub-system is being replaced and a decision is made to do this one at the same time. The data is vital for determining reliability statistics.
6. In some organisations a PM code structure is used. This can be very helpful but would benefit from a revision which allowed it to account for the difference between major rebuild, work resulting from maintenance tactics (e.g. fixed interval replacements, condition monitoring, and inspections), work generated by the operators or maintainers that is not part of a maintenance tactic and breakdowns. In addition, more accurate recording of PM codes is needed to ensure the correct code is being used.
7. Create separate work orders for each sub-system and maintainable item component that is changed out during major PM work such as a midlife.

If components that are changed out together need to be recorded (such as the final drives and differentials) these should be recorded separately so it is easy to see when each component was changed. This avoids the need to guess whether or not the differential has been changed out with the final drives or not.

8. Although not directly required for the reliability analysis, a correct value in the cost field as well as access to data on labour costs and downtime helps with deciphering the Short Text field. The cost helps in deducing the action in a work order (inspection, minor repair, major repair or change out) or the number of parts changed out if this information was not clearly given. It is possible to develop rules that make review of cost data and other entries so that spurious or obviously inaccurate entries are detected either prior to work order closure or as part of an end of month audit. It is possible to obtain the labour hours from elsewhere in the ERP system so this recommendation is about having standard (rather than special request) reports that include this information along with the other information described above.
9. Consideration should be given to how to create reports that can draw on data from maintenance plans, spares, maintenance contracts and equipment usage hours so that a more complete picture can be obtained for the asset. Currently these sit in separate places and have to be merged manually.

Future academic work should concentrate in two areas. The first is on improving data acquisition and measuring the maturity of the organisation in this area. This should build on the work of [16, 17]. The second is on developing measures for data quality specific to the types of issues highlighted here. Consistent measures allow for improvement projects to be assessed and also open the door to benchmarking across sites. Example target measures include:

1. Allocation of work to inappropriate levels in the asset hierarchy,
2. Inadequate or confusing identification of work,
3. Inappropriate use of Work order type and PM codes,
4. Incorrect use of maintenance tactic work order descriptors to breakdown work.

6 Conclusion

Getting useful information out of maintenance data may be challenging but it is not impossible. The insights gained from this project will assist organisations wishing to clean their own historical data (where necessary to support specific decisions) and improve data capture and coding. By providing details of the process and examples of issues, the authors hope to improve communication about these issues between the various parties involved, data collectors, data custodians and data users.

Acknowledgments The authors wish to thank CRC Mining for funding this project and to Peter Knights, research leader of CRC Mining's Equipment Management committee for his ongoing support.

References

1. (2012) Australian Bureau of Statistics: Industry Analysis. http://www.abs.gov.au/ausstats/. Accessed 12/3/2013
2. Lin S, Goa J, Koronios A (2006) The need for a data quality framework in asset management. In: Proceedings of the 1st Australasian workshop on information quality (AUSIQ), June 22–23
3. Batini C, Barone D, Federico C, Simone G (2011) A data quality methodology for heterogeneous data. Int J Database Manag Syst 3(1):60–79
4. Batini C, Cappiello C, Francalanci C, Maurino A (2009) Methodologies for data quality assessment and improvement. ACM Comput Surv 41(3):1–52
5. Borek A, Woodall P, Oberhofer M, Parlikad A (2011). A classification of data quality assessment methods. In: ICIQ 2011—proceedings of the 16th international conference on information quality
6. David L (2001) Data quality and business rules in practice. In: Enterprise knowledge management. Academic Press, San Diego
7. Lee YW (2004) Crafting rules: context-reflective data quality problem solving. J Manag Inf Syst 20(3):93–119
8. Lee YW, Strong DM (2004) Knowing-why about data processes and data quality. J Manag Inf Syst 20(3):13–39
9. Maydanchik A (2007) Data quality assessment. Technics Publications, New Jersey
10. Redman TC (2001) Data quality—the field guide. Digital Press, Oxford
11. Woodall P, Parlikad AK (2010) A hybrid approach to assessing data quality. In: Proceedings of the 2010 international conference on information quality
12. Hodkiewicz MR, Kelly P, Sikorska JZ, Gouws L (2006) A framework to assess data quality for reliability variables. World Congress on Engineering Asset Management (WCEAM), Gold Coast
13. Koronios A, Lin S (2004) Key issues in achieving data quality in Asset Management VETOMAC-3/ACSIM-2004 (Vibration engineering & technology of machinery, Asia-Pacific conference on system integrity & maintenance 2004, December 6–9) New Delhi
14. Lin S (2008) A data quality framework for engineering asset management. Aust J Mech Eng 5 (2):209–219
15. Haider A (ed) (2013) Information systems for engineering and infrastructure asset management. Springer, Berlin
16. Improving the Quality of Manually Acquired Data (2009) Applying the theory of planned behaviour to data quality. In: ICOMS 2009, Sydney, Australia
17. Unsworth K, Adriasola E, Johnston-Billings A, Dmitrieva A, Hodkiewicz M (2011) Goal hierarchy: improving asset data quality by improving motivation. Reliab Eng Syst Safety 96 (11):1474–1481
18. Murphy GD (2009) Building bridges and solving Rubiks Cubes: tribalism in engineering and technical environments. In: ICOMS 2009, Sydney, Australia
19. Hodkiewicz M, West G, Bartlett N, apsall S (2012) Asset management competency development. In: Lloyd C (ed) International case studies in asset management. ICE, Thomas Telford Press, London

The Application of Works Programme Management

Eight World Congress Engineering Asset Management Conference Paper 2013

Huazhuo Lin (Ling) and Roger Oed

Abstract Successful Works Programme Management (WPM) improves return on investment. Achieving successful works programme management is a challenge. It requires the right understanding of how to achieve the desired business outcomes. One size does not fit all, but instead the programme must be tailored to fit each specific time and place. The purpose of this chapter is to show how WEL Networks Ltd (WEL), as an electricity distribution utility, successfully implemented its WPM. WPM follows the model of; plan, do, act and check. This chapter covers the effective development of a 10 year asset management plan, and the selection of the best delivery model for effective and efficient delivery of the approved works programme. The chapter critically analyses work management models. In particular it identifies the benefits claimed for a WPM and some of the reported difficulties with implementation. It then considers the historical situation within WEL and shows how some of the reported problems were to be found within the WEL's history. It details the steps that WEL has taken in developing and implementing the WPM. It discusses the objectives and critical success factors in the implementation of WPM. The mechanisms for determining the effectiveness of the WPM implementation are described. The chapter concludes with lessons learned of how the WPM has been successfully applied to improve the effectiveness and efficiency of the development and delivery of the annual works programme in relation to the approved 10 year Asset Management Plan.

Eight world congress engineering asset management conference paper 2013.

H. Lin (Ling) (✉) · R. Oed
WEL Networks Limited, 114 Maui Street, Hamilton, New Zealand
e-mail: huazhuo.lin@wel.co.nz

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_31

1 Introduction

The purpose of this chapter is to share with other industry participants, the lessons learned in developing Works Programme Management (WPM) and its implementation within WEL Networks Ltd (WEL)—a New Zealand electricity distribution company.

2 Background

2.1 WEL Networks Ltd—the Company

WEL is the provider of electricity infrastructure, for the distribution of energy to over 85,000 homes, businesses and organisations throughout the Waikato region in New Zealand. The network includes more than 5,200 km of lines and has an annual throughput of 1,200 GWh. WEL has assets totalling in excess of NZ $550 million. The company employs 239 staff based at WEL's premises in Te Rapa, Hamilton. Approximately half are field staff and the other half are engineers and administrative staff.

WEL is locally owned. The company has one shareholder, the WEL Energy Trust. The capital beneficiaries are the region's local councils; Hamilton City Council, Waikato District Council and Waipa District Council.

2.2 Literature Review

Competitive advantage is key to the strategy of any company. The literature contains a number of strategic models for identifying and developing strategic advantage. Porter [7] developed the Industrial Organisation (I/O) perspective that considers the wider competitive environment (the market), which has proven value [5]. At the other extreme is the resource based view (RBV), which analyses the internal capabilities of a company and its resources which should be organised to gain a competitive advantage [3]. WEL is classed as a natural monopoly with virtually no external competition, so RBV is more relevant to WEL.

Porter [8] developed a value chain model for analysing the effectiveness with which a company manages its internal resources. This is most suited to a production line type company [6]. Stabell and Fjeldstad [10] refined RBV further and identified two other distinct types of company: consulting type companies and networking companies. The latter classification is particularly relevant to WEL. The different companies organise their resources in different ways [6]. The network type company continually needs to invest in its network to ensure it returns maximum value.

WEL invests about 10 % of total value of the network annually [13]. How should the internal resources be managed to optimise the investment?

Planning is critical. Extensive literature exists on optimising resources, but many papers focus on the re-ordering of stock, with a very formal methodology suitable for enterprise resource planning systems (ERP) [1, 2, 9, 11]. The main issue addressed in this chapter is not the management of hardware resources, but the scheduling of field staff to ensure the annual programme of work is completed. Toptal and Sabuncuoglu [12] describe distributed scheduling. WEL has adopted this model, which is described later in the chapter.

As stated above WEL invests about 10 % of its network value in upgrading and maintaining the network. About 70 % of that figure is planned work over a 10 year horizon. The other 30 % is unplanned and therefore uncertain. Part of it is a budget allocation for faults and the other part is an allocation for customer driven jobs, which can range from very small projects to major industrial installations. Feng et al. [4] describe resource scheduling for uncertain demand, but once again it is more applicable to the ordering function of an ERP system than managing staff resources. This chapter attempts to describe a practical approach.

2.3 Asset Management

It is imperative that a detailed Asset Management Plan (AMP) is in place given the long life of utility assets. By listening to the customers, by benchmarking with other New Zealand lines companies and by learning about trends overseas, WEL has established a number of objectives for the management of assets and these are detailed in the AMP. The primary objective is to optimise the performance of the network assets throughout their life cycle. Optimised performance covers safety, reliability, risk and return on investment.

Safety is paramount at WEL due to the dangerous nature of electricity and the risks it poses to staff and the general public. Isolating faults is a principle means of protecting people and plant. However, the AMP describes a move to ground fault neutralisation, which simultaneously maintains network safety and improves reliability, but at a higher capital cost. Thus all three aspects of safety, reliability and cost are interrelated. Hence each must be carefully balanced against the other to achieve optimum performance across all asset classes and across the entire life cycle of each asset.

The key inputs to the AMP are company strategic goals network performance, customer expectations, load forecasts and the outputs from the internal risk management process. Urgent safety issues are rectified immediately so are outside the AMP. Where a more strategic perspective is required the issues identified through the risk management process are addressed within the AMP. A network optimised for safety is one that has no safety related incidents. WEL has an uncompromising policy regarding safety; there are no budget constraints when correcting a safety problem. By comparison the optimisation of reliability is budget constrained. Investment in

reliability is optimised through cost, customer feedback, load forecasts and current network performance. Load forecasts are also very important for determining where the capacity of the network will need upgrading in the future. With a typical five year lead time for key infrastructure assets, the load forecasts are a critical factor in these investment decisions. The output of the AMP is a list of projects and budget allocations for the next ten years. Having a plan is fine, but unless it can be implemented it is of very little practical benefit. The next section discusses how to effectively and efficiently deliver the projects defined in the AMP.

3 History of Works Delivery Models

3.1 Overview

Figure 1 shows a timeline of when various works delivery models were adopted by WEL. Prior to 1992 WEL employed its own field staff. About this time ownership of the company changed and a decision was made to outsource the field staff. WEL was the first utility to do this in New Zealand. A surplus of labour existed at this time and a cultural change was needed within the company. Competition did bring about the desired cultural change. However, after ten years had elapsed problems with this model emerged. Managing the large number of comparatively small contractors was expensive. Further, since the companies were comparatively small they did not have the resources to invest in safety, plant or training. For these reasons it was decided to move to an alliance contract model. A contractor was appointed and one of the main benefits of the alliance was WEL learnt how to improve its safety practices.

The alliance contract ran for about 4 years before the environment changed again. Overseas governments recognised that their policies had led to major under

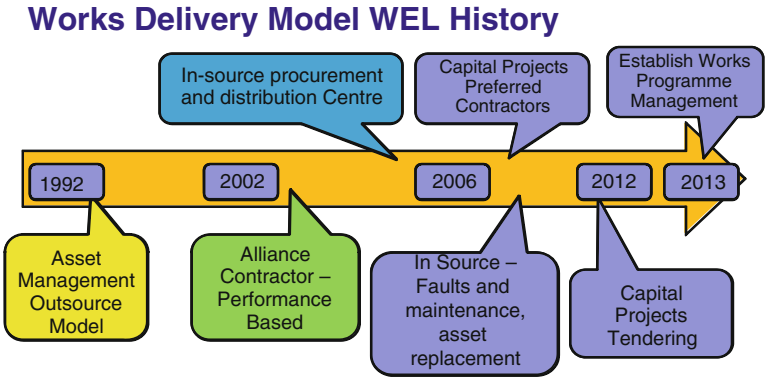


Fig. 1 A time line showing when the different models were adopted

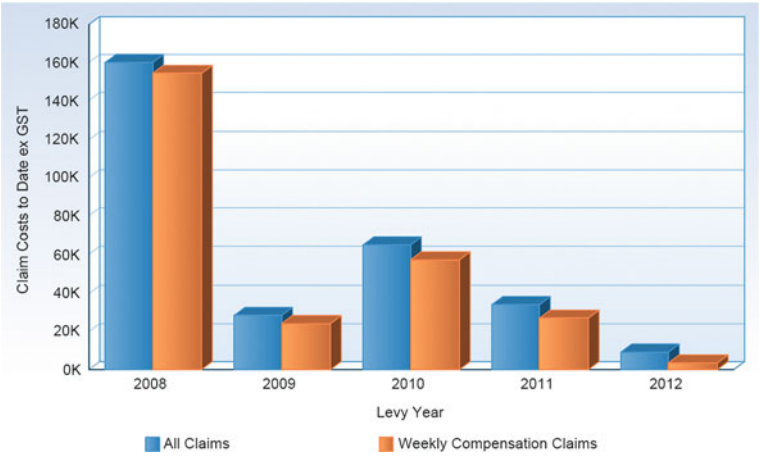


Fig. 2 Improvement in cost savings due to improved safety record

investment in their electrical networks. Suddenly, and in many countries across the world, utilities were spending large sums of money on network upgrades. Resources became scarce. No longer was there a surplus of field staff and competition had achieved the desired cultural change. WEL too ramped up its expenditure on the network at this time. However, it now faced the prospect of losing large numbers of qualified tradesmen overseas. A staff survey revealed they preferred working for the asset owning company rather than a contractor because it gave them better job security. The chosen solution was to terminate the alliance contract and in-source the field staff. This happened around 2006.

The main benefit from the alliance contractor model was a significant improvement in safety, which has continued into the present. A Health, Safety and Compliance team was set up to manage continued improvements. The overall public and staff safety has improved significantly.

The Total Incident Rate (TIR) is measured per 200,000 man hours. TIR includes both Lost Time Injuries (LTIs) and Medical Treatment Injuries (MTIs). This is done to bring focus and attention to MTIs which are one step below the LTIs in the injury pyramid. This improvement has led to a significant reduction in government levies as illustrated in Fig. 2.

3.2 Experience with the In-Sourced Model

Since in-sourcing the delivery function, WEL has better control over developing incentives and a culture designed to retain its workforce. In addition it has been able to exert tighter control over safety, quality and costs. Safety is discussed above. Quality has improved through the improvement of construction standards documentation and

the appointment of a field auditor. The improvement is measured with two indicators "Percentage Rework" and "Percentage Compliance" (to the WEL construction standard). Costs are controlled through two indicators, "Billability" and "Productivity". If used in isolation each indicator would drive behaviour to undesirable opposite extremes, but together they focus behaviour to deliver cost effective installations. Another area of improvement is more timely collection of as-built construction data with greater accuracy. The target accuracy is 100 %. Over recent years the consistency has dramatically improved from 40 to 90 %.

The most significant problem at present is the inability to deliver all the planned work in one financial year. Over the last 4 years, the average carry over was about $3 m or 10 % of the annual budget. The next sections investigate the cause of this and solutions for it.

4 Work Programme Management Implementation

4.1 Understanding How the Production Line Works

The delivery of the AMP can be modelled as a production line. The "raw materials" are the key elements of the asset management plan: asset replacement planning, customer jobs, network planning and the maintenance programme. The outputs from network planning, customer jobs and asset replacement planning all have to pass through the network design process. Thus the network design process is a constraint on the overall process. Further, in order to achieve a good "product" at the end, the work must be properly co-ordinated through WPM. A key function of WPM is how to optimally apportion work across both the internal and external resources. Figure 3 shows this to be the main bottleneck in the process, therefore it is a critical success factor for the whole process.

When the above model was applied within WEL analysis identified the lack of Works Programme Management as the reason for not completing all the planned

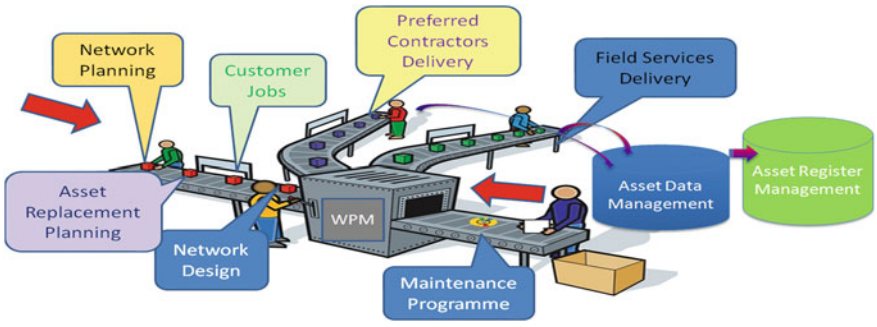


Fig. 3 Works programme management production line

projects within the required times. Poor planning impacts work scheduling in several ways:

- Poor definition of future resource requirements, leading to reactive or delayed responses to spikes in work load
- Having to accept quotes from contractors without the required commercial rigour, which leads to poor contracting strategy development and increased costs
- No firm work plan, resulting in short term scheduling, leading to poor efficiencies
- Over resourcing on jobs
- Inefficient inventory management.

Effective resource planning is only possible if the engineering designs are completed early enough. The key issues from this list are discussed in the following sections.

The results of the detailed process review of the works delivery processes are shown in the Fig. 4 below. An important aspect is the division of labour between the Asset Management function and Operations. Previously the planning and scheduling functions were co-located in the operations team. However, the Asset Management function is responsible for the longer term decisions and engineering design, and is also responsible for financial control of the work streams including initial budgetary approval. So long term resource planning was moved to the Asset Management function. Scheduling is typically performed over a three month time horizon, so is still part of Operations.

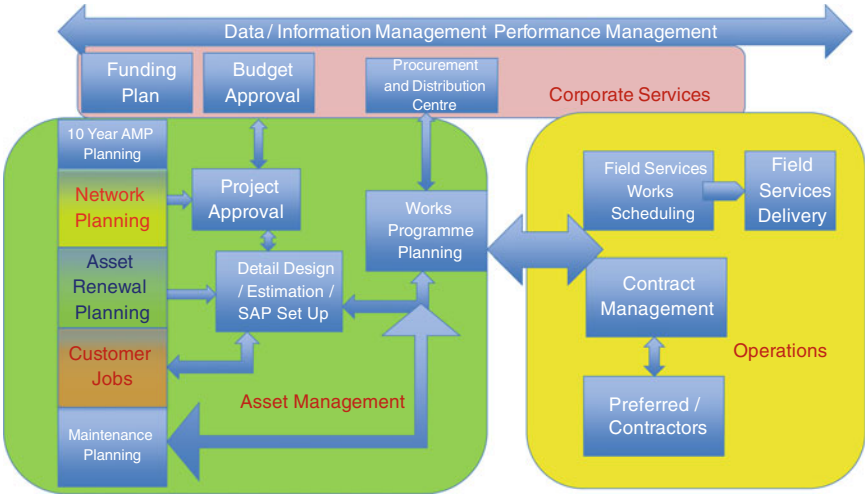


Fig. 4 Detailed analysis of works programme management

Works Programme Planning is therefore pivotal within Works Delivery processes and needs to be located within the Asset Management function. This is discussed in more detail in the next section.

4.2 Establishing the Right Position in the Right Group

The purpose of the Works Planning and Scheduling Process (WPASP) is to provide clear direction that the planning and scheduling groups can follow. By creating clear accountabilities within the division and establishing clear boundaries for the work that should be included it was anticipated that efficiencies could be gained, as well as improved working relationships across different groups and divisions. The anticipated strategic benefits include:

- Improved relationship between divisions
- Improved accuracy of information across all divisions
- Improved efficiency of the Works Delivery group.

The WPASP is focused on providing long term planning functions for strategic decision making, as well as providing a balanced three monthly schedule of work. The group must also balance all external workforce requirements. The outputs of this process are:

- Complete job packs, containing the design pack, maintenance pack, or action register with allocated resources
- Updated Global High Level Plan (18 month rolling)
- Updated Master Plan and Schedule (three month rolling).

It was decided to establish a works programme manager position within the Asset Management group rather than in Field Service group. This position is responsible for producing a resource plan to implement the approved 18 month rolling work plan. The position is also responsible for producing an approved AMP that includes the maintenance plan in order to:

- Identify forward resource requirements for the in-house work force capacity and competency development
- Identify forward resource requirements for preferred contractors to ensure proper contracting strategy management
- Identify projects to be tendered for a competitive price or benchmarking purpose
- Identify long lead time material delivery issues for the Procurement Manager to develop proper procurement strategy to mitigate risks, allowing just in time delivery of high value capital purchases
- Communicate the long term Resource Plan to key stakeholders and update it based on feedback and obtain approval for on-going, regular monitoring and updates (monthly or quarterly), and provide coordinated feedback to key stakeholders

- Monitor the overall capital and maintenance programme delivery performance and take corrective actions to achieve targets.

The next section discusses the recent implementation of the Works Programme Planning process.

4.3 Implementation

The role of Works Programme Manager was established. The priority tasks included the following:

- Capital project data was collected from project managers and budgets. This information was transferred to MS Excel to allow a resource "snapshot" of the information "today"
- All maintenance activities (including faults) budgeted in this financial year from the Maintenance Strategy team were applied across the various Field Service crafts
- The Field Services resource availability (actual head count by craft) plus overtime was used to obtain a visual "supply and demand" resource comparison.

Figure 5 shows the required resources (light blue bars), which is compared with the available in-house Field Services resources. The in-house resources were shown to be significantly below that required to meet the demand.

In order to fairly allocate work to in-house Field Services, staff workshops were held and established the following:

- Field Services will be responsible for faults, maintenance programme (vegetation is a separate programme), plus as much as possible of the asset replacement projects, and a selection of capital projects
- Capital project selection (including customer jobs) for Field Services will be based on resource availability in consideration with seasonal weather patterns, combining the maintenance programme and asset replacement projects

Fig. 5 Total planned work compared against operations capacity

- Skilled staff retention would be enhanced by allocating suitable and challenging work in house. E.g. Technician staff or other crafts carry out suitable portions of out sourced projects.

Where Field Services have insufficient resources to deliver the work programme it must be contracted out in order to complete all the work by the end of the budget period. In the past this has happened on an ad hoc basis often near the end of the budget period. At this time the contractors often had too much work, so WEL was charged a premium by them or simply could not engage anyone to complete the work. Improved planning will allow WEL to enter into a preferred contractor agreement and provide a minimum guaranteed workload for them at the beginning of the financial year. This has two key benefits. Firstly, WEL can obtain competitive prices for its work and secondly, the works programme can be completed within the budget period without any carry over to the next year.

After completing the high level analysis, a detailed analysis of resource requirements by craft was performed in order to efficiently allocate work to Field Services. The analysis showed where adequate resources are available and where there are currently shortages. An example is shown Fig. 6.

It shows a significant shortage of cable jointers. This is a major impediment to the works delivery programme. These people are highly skilled, rare and take a long time to train. There is therefore potentially a long lead time required to increase the internal resources. This is an area that may have to be contracted out, at least in the short term.

Effective resource planning is only possible if the engineering designs are completed early enough. Early completion of designs means early identification of resource and material constraints is possible. This provides sufficient time to mitigate these risks before they impact on the performance of WEL. Other benefits include aggregation of material requirements from various projects, which creates the possibility of discounts for bulk purchases, as well as optimising scheduling according to the geographic location of projects. As a result of the constraints in the Design Team, approval has been granted to employ two more designers. It may be necessary to hire external consultants to assist in levelling out some of the workload of the Design Team. A software tool can aid scheduling and provide transparent communication across all teams within the company.

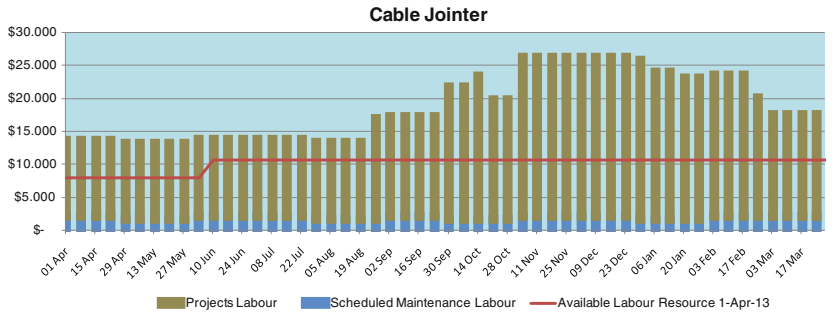


Fig. 6 Cable Jointer resource shortage

4.4 Lessons Learned

The key lessons learned after 20 years' experience is:

- No one model is perfect, but companies must choose the right model at the right time in the right environment
- The model of letting work out to a large number of smaller contracting companies can be useful for changing the culture of the industry, but it has a comparatively short life span. The limitations are due to the inability of smaller contracting companies to invest in safety, plant and training and maintain technical standards
- An alliance model overcomes many of these limitations. It can provide effective knowledge transfer. However, the asset owning company still lacks control over staff retention, as well as safety and quality to some extent. Where staff shortages exist or control of safety, quality or costs become an issue, then an in-house model is preferred. Risk management and retention of corporate intellectual property may also be at risk with this model
- The in-house model is a good model if staff retention is an issue. The drivers foster a pride in the network and personal ownership of the network. The difficulties with the in-sourcing model are maintaining a productivity orientated culture and ensuring that the in-house cost structure is competitive and comparable to the wider market costs
- Long term planning is essential for successful implementation of a works delivery programme. The design function must be sufficiently resourced to enable it to produce designs up to a year before they are needed for construction. The detailed designs are needed for detailed resource planning
- The recent implementation of works programme management has shown that a Works Programme Manager is better located in the Asset Management function and the schedulers within the Operations group
- Efficient delivery is achieved through earlier and more comprehensive planning
- Communication across the groups is vital. A software tool is important for aiding that visibility across groups.

5 Conclusion

The various works delivery models have been described and critically analysed. The key findings are as follows:

- Companies must understand the operating environment to understand which operating model works best for their environment. The business environment is dynamic, so the validity of an existing operating model must be constantly reviewed and modified where necessary

- WEL has implemented this principle for over 20 years, which initially has resulted in out-sourcing, then moving to an alliance contract before moving back to in-sourcing
- Far greater control has been gained through the in-sourcing option resulting in greater stability of wage rates and lower costs. At the same time, the sense of ownership has resulted in the improvement of both safety standards and quality
- In order to maintain cost effectiveness, carefully designed measures must be devised to ensure productivity remains high, yet cost effective.

A finding common to all models is that long term asset management and related resource planning are essential for successful implementation of a works delivery programme. In order to enable good resource planning detailed designs must be produced up to a year prior to construction. For this to happen the design function must be sufficiently resourced, so in order to ensure adequate delivery resources there must be adequate planning and design resources located in the Asset Management function of the organisation.

References

1. Ahuja V, Thiruvengadam V (2004) Project scheduling and monitoring: current research status. Constr Innov 4:19–31
2. Ait_Kadi D, Menye JB, Kane H (2012) Resources assignment model in maintenance activities scheduling. Int J Prod Res 49(22):6677–6689
3. Barney J (1991) Firm resources and sustained competitive advantage. J Manag 17(1):99
4. Feng K, Rao US, Raturi A (2011) Setting planned orders in master production scheduling under demand uncertainty. Int J Prod Res 49(13):4007–4025
5. Korsaa CR (2010) Integrating business models and strategy for sustained competitive advantage—a case study of Ryanair; MSc in Economics and Business Administration. Department of Marketing, Copenhagen Business School
6. Othman R, Sheeham NT (2011) Value creation logics and resource management: a review. J Strategy Manag 4(1):5
7. Porter ME (1985) Competitive advantage: creating and sustaining superior performance. Free Press, Collier Macmillan, New York, London
8. Porter ME (1998) Competitive advantage: creating and sustaining superior performance. Free Press, New York
9. Segerstedt A (2006) Master production scheduling and a comparison of material requirements planning and cover-time planning. Int J Prod Res 44(18–19):3585–3606
10. Stabell DG, Fjeldstad OD (1998) Configuring value for competitive advantage: on chains, shops and networks. Strategic Manag J 19:413–437
11. Talbot FB (1982) Resource-constrained project scheduling with time-resource tradeoffs: the nopreemptive case. Manag Sci 28(10):1197–1210
12. Toptal A, Sabuncuoglu I (2010) Distributed scheduling: a review of concepts and applications. Int J Prod Res 48(18):5235–5262
13. WEL Networks Ltd (2013) Asset management plan for period 1 April 2013 to March 2023. www.wel.co.nz/Corporate-And-Community/Regulatory-Disclosuress/Asset-Management-Plans/

A Performance Degradation Interval Prediction Method Based on Support Vector Machine and Fuzzy Information Granulation

Fuqiang Sun, Xiaoyang Li and Tongmin Jiang

Abstract To predict the trend and interval of the product performance degradation, a combination approach of fuzzy information granulation (FIG) and support vector machine (SVM) is proposed. Firstly, to make interval prediction of performance degradation and reduce prediction error, the monitoring performance degradation data is divided into several segments in accordance with the actual needs, and the fuzzy information granulation method is used to describe the information of each data segment by the concept of information granule. Then, the support vector machine is applied in the modelling of the fuzzy information granules data. Finally, the proposed FIG–SVM method is applied in degradation assessment of a microwave product, and the result shows that the method is feasible and is effective in improving the modelling precision.

1 Introduction

With the rapid growth of technology and market competition, there are some new characteristics of the modern products development direction, such as complexity, high-speed and intelligence. And also the products must face more and more harsh running condition. Once faults occur in these products, it will affect the production efficiency and induce economic loss, or even bring up personnel casualty and cause serious social effects. Therefore, it is necessary and important to monitor product's

This work was supported by the Fundamental Research Funds for the Central Universities.

F. Sun (✉) · X. Li
Science and Technology on Reliability and Environmental Engineering Laboratory, Beihang University, No. 37 Xueyuan Road, 100191 Beijing, People's Republic of China
e-mail: sunfuqiang@buaa.edu.cn

T. Jiang
School of Reliability and Systems Engineering, Beihang University, Beijing, People's Republic of China

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_32

operation state and diagnose their faults in order to improve safety, allow predictive maintenance and shorten significantly the associated out of service time.

Generally, it is a continuous performance degradation process from the product initial state to the final failure. If the product performance degradation degree and trend can be obtained by monitoring the characteristic parameters, it would be possible to make credible maintenance schedule and prevent the urgent broken. Product performance degeneration assessment and trend prediction is proposed based on the above idea, which is the key technology of the Prognostic and Health Management (PHM) and the Intelligent Maintenance System (IMS) [1].

The essence of degradation assessment and trend prediction is the pattern recognition of product operation state. Most of the existing degradation assessment and prediction methods can be divided into two main categories: physics-based models and data-driven models [2]. Physics-based models typically involve building technically comprehensive mathematical models to describe the physics of the system and failure mechanism. However, for most industry applications, physics-based models might not be the most practical solution since the system composition and failure mechanism are too complex. Data-driven approaches attempt to obtain models directly from routinely collected performance degradation data instead of building models based on comprehensive system physics and human expertise. In recent years, the commonly used data-driven modeling methods including ARMA models, Artificial Neural Networks (ANN), Kalman filtering, grey model, and hidden Markov model (HMM). Among these methods, the ANN method is more widely adopted. However, the ANN method is based on the empirical risk minimization (ERM) principle. As a result, the ANN method might run into the over-fitting or under-fitting problem and get the bad generalization performance.

Recently, the support vector machine (SVM), a novel learning machine based on statistical learning theory (SLT), was developed by Vapnik and his co-workers in 1995 at the AT&T Bell Laboratories [3]. The SVM implements the structural risk minimization (SRM) principle, which makes the SVM overcome the above drawbacks of the ANN method and have a preferable generalization capability. Besides that, the SVM algorithm is equivalent to solving a quadratic programming problem, which can ensure the solution to be global optimization. Just for these advantages, the SVM method has been successfully applied in various fields including finance, control, engineering, etc. In [4, 5], SVM regression is respectively applied to machine condition trend prediction based on vibration signals. In [6], the SVM is utilized in the drift modelling of the dynamically tuned gyroscope (DTG).

However, the SVM multi-step prediction error has the increasing regularity with the increase of the number of prediction steps [7]. In order to make long-term prediction for the product performance state and reduce prediction error, this chapter divides the degradation data into several data segments, and uses a certain description to replace the information of each data segment. On this basis, the multi-step prediction model for the description of each segment data could be established. So, the key question is how to accurately describe the information of each data segment. Moreover, in many cases, the purpose of the performance

degradation assessment and prediction is not only to accurately predict the performance status at future point in time, but also to predict the trend and interval of the products performance state in the next period of time (as planned downtime).

Fuzzy information granulation (FIG) method provides an effective way to solve these problems. Information granule has outstanding advantages in terms of processing vague, incomplete, imprecise and uncertain data mining. It can ignore the irrelevant details of the interval data, and the characteristics of the interval can be well characterized at the same time [8].

Therefore, an integration degradation assessment and trend prediction method of fuzzy information granulation and support vector machine is proposed in this chapter. Fuzzy information granulation method adopts the concept of information granule to describe each data segment. By modeling the fuzzy granules using SVM, the product performance state changing space in the future could be predicted. The proposed FIG-SVM method is expected to more accurately identify the behaviour of the products performance degradation and to provide the basis for maintenance decision-making better.

2 Support Vector Machine for Regression

Support Vector Machine (SVM) was developed by V. N. Vapnik and his co-workers [3, 9], which is a novel learning machine based on statistical learning theory. SVM minimizes the true risk by principle of Structure Risk Minimization (SRM). In 1997, with the introduction of Vapnik's ε-insensitive loss function, SVM was successfully utilized to solve the regression problems. The basic idea the SVM for regression is to map the input data into a high-dimensional feature space by a nonlinear mapping Φ and to perform the linear regression in this space, as shown in Fig. 1.

The theory of SVM for regression is briefly introduced as follows [6, 7].

Given a set of training data $\{(x_i, y_i), \ldots (x_l, y_l)\}$, $i = 1, 2, \ldots, l$, where $\boldsymbol{x}_i \in \mathbf{R}^n$ denote input patterns, $y_i \in \mathbf{R}$ are the targets and l is the total number of training samples. In SVM for regression, the goal is to find a function $f(x)$, i.e. an optimal hyperplane, which has at most ε deviation from the actually obtained targets y_i for all the training data and is as flat as possible. The form of functions is denoted as

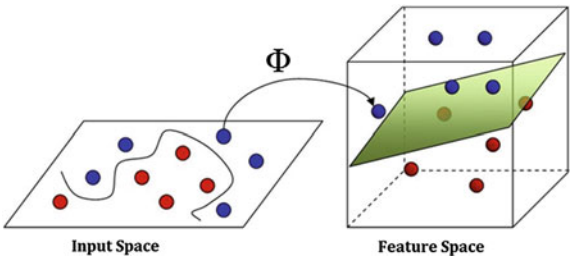


Fig. 1 The basic idea of the SVM

$$y = f(x) = \langle \omega, \Phi(x) \rangle + b \quad \text{with} \quad \Phi : R^n \to F, \omega \in F \tag{1}$$

where $\Phi(\cdot)$ is a nonlinear mapping by which the input data x is mapped into a high dimensional space $\boldsymbol{F}$, and $\langle \cdot, \cdot \rangle$ denotes the dot product in space $\boldsymbol{F}$. The unknown variables ω and b are estimated by minimizing the regularized risk function $R_{\text{reg}}[f]$

$$R_{\text{reg}}[f] = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l} \zeta(f(x_i) - y_i) \tag{2}$$

$$\zeta(f(x) - y) = \max\{0, |y - f(x)| - \varepsilon\} \tag{3}$$

where C is a constant which determining the trade-off between the flatness of the regularized term ($\|\omega\|^2/2$) and the empirical error. $\zeta(\cdot)$ is the ε-insensitive loss function, and e is the tube size.

By introducing the non-negative slack variables $\xi_i^{(*)}$, Eq. (2) is transformed into the following convex constrained optimization problem:

$$\text{Minimize } \Gamma(\omega, \xi, \xi^*) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l} \zeta(\xi_i + \xi_i^*) \tag{4}$$

$$\begin{aligned} \text{Subject to} \quad & \langle \omega, \Phi(x_i) \rangle + b - y_i \le \varepsilon + \xi_i \\ & y_i - \langle \omega, \Phi(x_i) \rangle - b \le \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \ge 0, \quad i = 1, 2, \ldots, l. \end{aligned} \tag{5}$$

According to Wolfe's Dual Theorem and the saddle-point condition, the dual optimization problem of the Eq. (4) is obtained as the following form:

$$\begin{aligned} \text{Minimize } W(\alpha_i^{(*)}) = & \varepsilon \sum_{i=1}^{l} (\alpha_i^* + \alpha_i) - \sum_{i=1}^{l} y_i(\alpha_i^* - \alpha_i) \\ & + \frac{1}{2}\sum_{i,j=1}^{l} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\langle \Phi(x_i), \Phi(x_j) \rangle \end{aligned} \tag{6}$$

$$\begin{aligned} \text{Subject to } & \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0 \\ & 0 \le \alpha_i^{(*)} \le C, \ i = 1, 2, \ldots, l. \end{aligned} \tag{7}$$

$$\text{With} \quad \omega = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i)\Phi(x_i) \tag{8}$$

where $\alpha_i^{(*)}$ are the nonnegative Lagrange multipliers that can be obtained by solving the convex quadratic programming problem stated above. Finally, by exploiting the

Karush–Kuhn–Tucker (KKT) conditions, the decision function given by Eq. (1) gets the following form:

$$f(x) = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i)\langle \Phi(x_i), \Phi(x) \rangle + b \tag{9}$$

In order to get the dot product in the feature space simply and avoid the curse of dimensionality, the kernel function $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ is introduced, which satisfies the Mercer condition. The most commonly used kernel functions are:

(a) The polynomial kernel $k(x, x') = (\langle x, x' \rangle + c)^p,\ p \in N, c \geq 0$

(b) The Radial Basis Function (RBF) kernel $k(x, x') = \exp\left(-\|x - x'\|^2 \Big/ (2\sigma^2)\right)$

(c) The Sigmoid kernel $k(x, x') = \tanh(b\langle x, x' \rangle + \theta)$

3 Theory of Fuzzy Information Granulation

3.1 Information Granulation

The concept of information granule was first proposed by L.A. Zadeh in 1979 [10, 11]. Informally, the information granules are collection of entities, usually originating at the numeric level, that are arranged together due to their similarity, functional adjacency, indistinguishability, coherency or alike. The information granulation (IG) is a process of the construction of information granules.

There are mainly three models of IG: model based on fuzzy set theory; model based on rough set theory; model based on quotient space theory [12]. This chapter emphasizes on model based on fuzzy set theory without concerning another two models.

Fuzzy set theory was proposed by L.A. Zadeh in 1960s. Based on fuzzy set, Zadeh gave a profile of data granule in fuzzy information granulation (FIG) problem as follows

$$g = (x \text{ is } G) \text{ is } \lambda \tag{10}$$

where x is variable of universe U, G is a fuzzy subset of U, the membership function of which is $\boldsymbol{\mu}_G$, λ denotes the probability.

3.2 FIG Method for Time Series

A. Bargiela and W. Pedrycz introduced the theory of information granulation (IG) to time series analysis [13, 14, 15]. They presented there an optimized fuzzy information granulation method for building fuzzy granules.

FIG method of W. Pedrycz has two steps: firstly, the given sequence should be divided into several subsequences as the operation windows; Secondly, fuzzification processing in every window to generate the fuzzy set, i.e. fuzzy granule.

Given a sequence $\boldsymbol{X} = (x_1, x_2, \ldots, x_n)$, consider the single window problem, i.e. get a fuzzy granule P from $\boldsymbol{X}$. The fuzzy granule P is a fuzzy concept $\boldsymbol{G}$ which is a fuzzy set with universe $\boldsymbol{X}$. The process of FIG is to get the membership function A of $\boldsymbol{G}$, i.e. $A = \boldsymbol{\mu}_{\mathrm{G}}$. In the following statements, $\boldsymbol{G}$ can be substituted by fuzzy granule P without special announcement, i.e.

$$P = A(x),\ x \in \boldsymbol{X} \tag{11}$$

The main forms of fuzzy granules including: triangle, trapezoid, gauss and parabola, etc. The membership function of triangle fuzzy granule in this chapter is:

$$A(x, a, m, b) = \begin{cases} 0, & x < a \\ \frac{x-a}{m-a}, & a \le x \le m \\ \frac{b-x}{b-m}, & m < x \le b \\ 0, & x > b \end{cases} \tag{12}$$

where a, m, and b are parameters of A . A graph of triangle fuzzy granule is shown in Fig. 2.

The basic ideas of constructing fuzzy granule are proposed by W. Pedrycz.

(i) Fuzzy granule should represent the original data reasonably, i.e. maximizing the sum of membership function A;

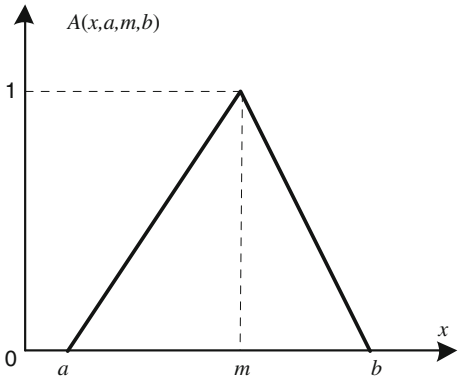


Fig. 2 The membership function of triangle fuzzy granule

(ii) Fuzzy granule should have some particularity, i.e. minimizing measure(supp-(*A*)), which is called the support measure of *A*.

Based on the ideas, a function of *A* is constructed to find a balance of ideas (i) and (ii):

$$Q(A) = \frac{M(A)}{N(A)} = \frac{\sum_{x \in X} A(x)}{\text{measure}(\text{supp}(A))} \tag{13}$$

where $M(A)$ satisfies idea (i) and $N(A)$ satisfies idea (ii), maximizing $Q(A)$ can satisfy the balance of ideas (i) and (ii).

4 The FIG–SVM Prediction Method

In practice, whether the product is running in good condition at a future period of time is concerned by enterprises, which is an important basis for developing production plans. Therefore, the FIG-SVM combined method is utilized to assess and predict the degradation trend and interval of product performance state in the future period of time (e.g. a week). The framework of FIG-SVM degradation interval prediction method is shown in Fig. 3.

Assume that a product has p performance characteristic parameters and the number of detections of each parameter is M. The observations matrix of product performance parameters is:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{M1} & x_{M2} & \cdots & x_{Mp} \end{bmatrix} \tag{14}$$

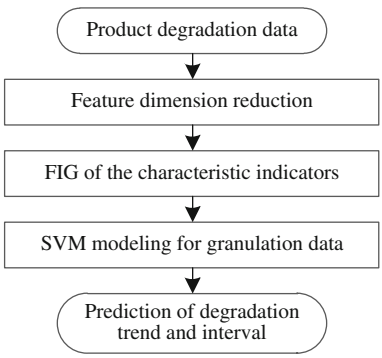


Fig. 3 Framework of the FIG-SVM prediction method

where, $x_{ij}, i = 1, 2, \ldots, M, j = 1, 2, \ldots, p$ denotes the observed value of the i-th detection of the j-th performance characteristic parameter.

Products performance degradation assessment is actually a dynamic pattern recognition problem. To reduce the calculation complexity of the pattern classifier, it is necessary to carry out feature dimension reduction for the original data firstly. Principal component analysis (PCA) is one of the most commonly used statistical methods for feature extraction and dimension reduction by transforming the original features into a new set of uncorrelated features [16]. The obtained principal components can serve as the characteristic indicators of products performance state.

4.1 FIG Processing

To make interval prediction of performance degradation and reduce prediction error, the degradation data is divided into several segments, and FIG is used to describe the information of each data segment by fuzzy granules. The W. Pedrycz method of constructing fuzzy granules is utilized here. The FIG algorithm of the product performance characteristic indicators, taking the first principal component y_1 as example, is shown below.

I Let the operation window size is w, and divide the principal component y_1 into $[M/w]$ subsequences, denoted by Δy_n, n = 1,2,…, $[M/w]$.

II Construct the triangle fuzzy granule $P = (a, m, b)$ of a subsequence Δy_n. Firstly, determine the median of subsequence Δy_n as the core of fuzzy granule P, i.e. m. Then, determine the parameters a and b of fuzzy granule P by solving the optimization problem in Eq. (15).

$$Maximize\ Q(A) = \frac{\sum_{i=1}^{w} A(\Delta y_{ni})}{\text{measure}(\text{supp}(A))} = \frac{\sum_{i=1}^{w} A(\Delta y_{ni})}{b - a} \tag{15}$$

III Using the above methods, construct the triangle fuzzy granules $P_i = (a_i, m_i, b_i)$ i = 1, 2, …, $[M/w]$ of the $[M/w]$ subsequences of the principal component y_1, respectively. Denote

$$\begin{cases} \text{Low} = [a_1, a_2, \ldots, a_{[M/w]}]' \\ \text{R} = [m_1, m_2, \ldots, m_{[M/w]}]' \\ \text{Up} = [b_1, b_2, \ldots, b_{[M/w]}]' \end{cases} \tag{16}$$

where, Low, R, and Up respectively describe the minimum value, average level, and maximum value of principal component y_1 changing in the $[M/w]$ subsequences.

The other principal component data could conduct fuzzy information granulation processing by the same method.

4.2 SVM Modeling

The data phase space reconstruction is conducted in order to mine the data for more useful information, i.e. making a one-dimensional time series $X_N = \{x_1, x_2, \ldots, x_N\}$ transformed into the following matrix form

$$X_{re} = \begin{bmatrix} \vec{x}_{h+1} \\ \vec{x}_{h+2} \\ \cdots \\ \vec{x}_N \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_h \\ x_2 & x_3 & \cdots & x_{h+1} \\ \cdots & \cdots & \ddots & \cdots \\ x_{N-h} & x_{N-h+1} & \cdots & x_{N-1} \end{bmatrix}, \; Y_{re} = \begin{bmatrix} x_{h+1} \\ x_{h+2} \\ \cdots \\ x_N \end{bmatrix} \tag{17}$$

where, h is the prediction embedded order. The principle of Final Prediction Error (FPF), which can identify AR model order in time series analysis, is utilized to obtain the best prediction order h [17].

The SVM regression model is constructed with X_{re} as the input matrix and Y_{re} as the target vector.

$$x_i = f(\vec{x}_i) = f(\{x_{i-h}, x_{i-h+1}, \ldots, x_{i-2}, x_{i-1}\}) \tag{18}$$

The SVM regression model (18) is utilized for predicting the future trend. The l-steps SVM prediction model is as follow:

$$\hat{x}_{N+l} = f(\{x_{N-h+l}, x_{N-h+l+1}, \ldots, x_N, \hat{x}_{N+1}, \ldots, \hat{x}_{N+l}\}) \tag{19}$$

For granulation data Low, R, and Up of every principal component, the space reconstruction are conducted according to Eq. (17) firstly. Then, the SVM regression models are respectively constructed. Finally, the recursive prediction of the granulation data Low, R, and Up of every principal component is carried out based on the l-steps SVM prediction model in Eq. (19). Consequently, the change trend and intervals of product operation state can be obtained by this method.

5 Case Study

Taking a microwave product as the application object, the proposed FIG-SVM method is utilized to predict the change trend and intervals of its performance state. This product has 9 performance characteristic parameters, and we collected a total of 686 observations of every parameter by detecting once a day. The purpose of this case is to predict the change trend and intervals of product performance in the coming weeks. The former 448 daily degradation data are used in the FIG-SVM modeling, as shown in Fig. 4. And the remaining data will serve as the test data to verify the modeling accuracy.

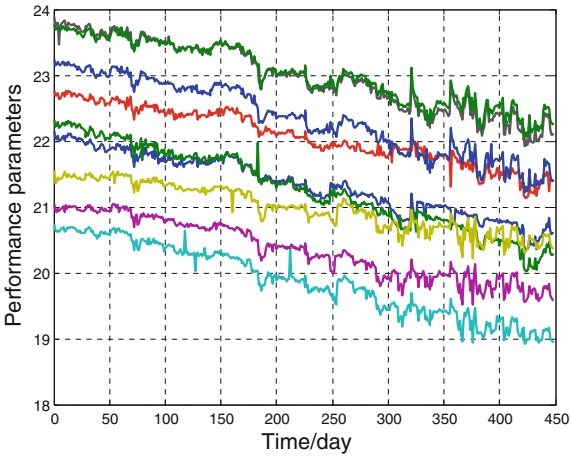


Fig. 4 The observation data of performance parameters

PCA is used to reduce the dimension of the original observation data. In this case, the cumulative contribution rate of the first principal component is already more than 90 %. Therefore, the product performance degradation characteristic indicator is the first principal component, as shown in Fig. 5.

Then, the W. Pedrycz FIG method is utilized to deal with the degradation characteristic indicator. The granulation window size is 7, i.e. the length the subsequence is 7, and the first principal component data is divided into 64 operation windows. Then, the fuzzy granules are constructed on every window, and the results are shown in Fig. 6.

Finally, according to the Eq. (19), the 34-step SVM prediction models for the granulation data Low, R, and Up are established. And the degradation trend and intervals of product performance state can be obtained by the recursive prediction method, as shown in Fig. 7.

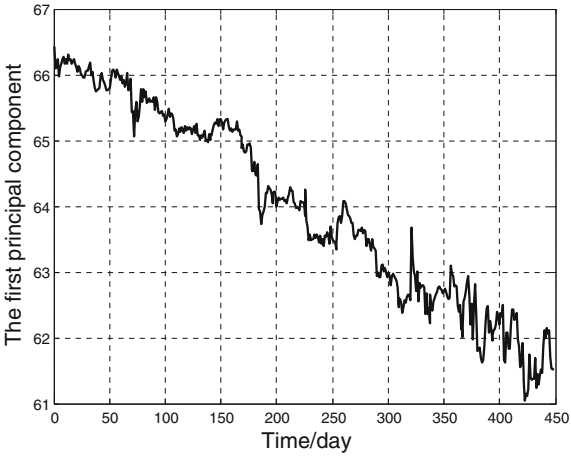


Fig. 5 The first principal component

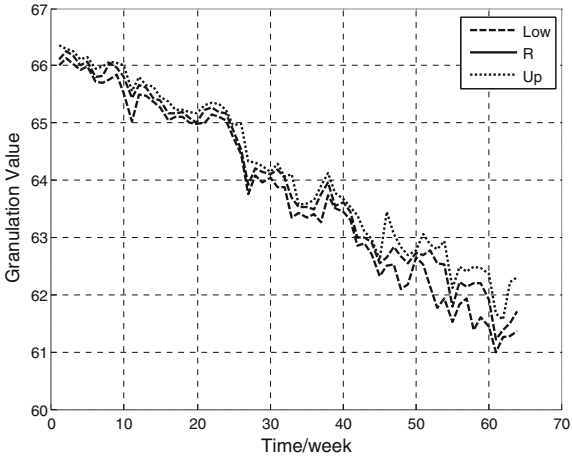


Fig. 6 The results of FIG processing

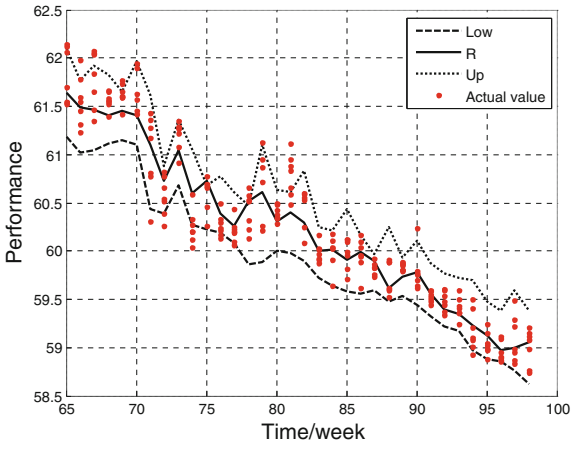


Fig. 7 The degradation trend and intervals of product performance state

It can be seen from Fig. 7 that the predicted result consistent with the test data, i.e. actual product performance status. And the proposed method is accurate and feasible. According to the prediction results, the changes of performance status in the coming weeks can be obtained, which can provide assistance for the maintenance decision.

6 Conclusion

In order to achieve the best utilization of the product, reduce maintenance costs and improve production efficiency, product performance degradation assessment theory and technology is researched in this chapter. A combination approach of the fuzzy

information granulation and support vector machine is proposed to construct the prediction model of products performance degradation. This approach quantitative describes the product running status and predicts the trend and intervals of product performance degradation. The case analysis showed that the proposed method can accurately describe the degree of performance degradation.

At present, the study of product performance degradation assessment and prediction is still in its infancy, and the proposed model has laid a foundation, but there are a lot of research needs to be further carried out.

References

1. Lee J, Ni J et al (2006) Intelligent prognostics tools and e-maintenance. Comput Ind 57 (6):476–489
2. Heng Aiwina, Zhang Sheng, Tan Andy CC, Mathew Joseph (2009) Rotating machinery prognostics: State of the art, challenges and opportunities. Mech Syst Sig Process 23:724–739
3. Vapnik VN (1995) The nature of statistical learning theory. Springer, New York
4. Lin H (2006) Support vector machines for regression and its application for prediction of machine degradation based on vibration signals. Master's thesis, University of Alberta
5. Li L-J, Zhang Z-S, He Z-J (2004) Research on condition trend prediction of mechanical equipment based on support vector machines. J Xi'an Jiaotong Univ 38(3):230–238
6. Guoping Xu, Tian Weifeng, Jin Zhihua (2006) An AGO–SVM drift modeling method for a dynamically tuned gyroscope. Meas Sci Technol 17(1):161–167
7. Guoping Xu, Tian Weifeng, Qian Li (2007) EMD- and SVM-based temperature drift modeling and compensation for a dynamically tuned gyroscope (DTG). Mech Syst Sig Process 21(8):3182–3188
8. Sun F, Li X, Jiang T (2011) A novel performance degradation interval prediction method based on support vector machine and fuzzy information granulation. China Patent No. ZL 201110203058.2
9. Smola AJ, Scholkopf B (2004) A tutorial on support vector regression. Stat Comput 14 (3):199–222
10. Zadeh LA (1979) Fuzzy Sets and Information Granularity. Advances in fuzzy set theory and applications. North-Holland, Amsterdam, pp 3–18
11. Zadeh LA (1997) Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst 90(2):111–127
12. Li Y, Yu F (2010) Optimized fuzzy information granulation based machine learning classification. In: Proceedings of seventh international conference on fuzzy systems and knowledge discovery (FSKD), pp 259–263
13. Pedrycz Witold (2005) Knowledge—based clustering—from data to information granules. Wiley, New York
14. Bargiela A, Pedrycz W (2003) Granular computing: an introduction. Kluwer Academic Publishers, Dordrecht
15. Dong K (2005) Time series information granulation and granulation-based cluster analysis. Master's thesis, Beijing Normal University
16. Zheng Y, Sun G (2013) A novel power system reliability predicting model based on PCA and RVM. Math Probl Eng Vol 2013:1–6
17. Lei G (2009) Study on Kernel pattern analysis methods based rotating machinery performance degradation assessment technique. PhD's thesis, Shanghai Jiao Tong University

Maintenance Solutions for Cost-Effective Production: A Case Study in a Paper Mill

Damjan Maletič, Viktor Lovrenčič, Matjaž Maletič, Basim Al-Najjar and Boštjan Gomišček

Abstract For companies, in order to stay competitive, it is necessary to continuously increase the effectiveness and efficiency of their production processes. Therefore the purpose of this chapter is to discuss the role of maintenance in achieving the competitive advantages using cost-effectiveness aspect of maintenance process. In this regard, the chapter illustrates/discusses the impact of mechanical and electrical failures on company's business on an example of a paper mill where processes are running 24/7. Thus, this paper presents the role of vibration-based maintenance (VBM) in enhancing the production and maintenance performance continuously and cost-effectively. Using empirical data collected from a paper mill case study, we found that company could avoid the profit losses even to a greater extent if it would improve the effectiveness of the VBM. With respect to the electrical causes of failures, a live working technique for improving the reliability and availability of the paper machine is proposed. Therefore, maintenance solution concerning the paper machine is suggested and discussed as well as potential benefits are highlighted. The results supported the notion that there is a positive association between the reduction of the unplanned stoppages and potential savings. The results have also shown that there is a need for more systematic approach, and a more holistic view of the maintenance function for establishing and running a cost-effective maintenance policy in the paper mill under consideration.

D. Maletič · M. Maletič · B. Gomišček (✉)
Faculty of Organizational Sciences, University of Maribor, Maribor, Slovenia
e-mail: bostjan.gomiscek@fov.uni-mb.si

D. Maletič
e-mail: damjan.maletic@fov.uni-mb.si

M. Maletič
e-mail: matjaz.maletic@fov.uni-mb.si

V. Lovrenčič
C&G D.O.O. Ljubljana, Ljubljana, Slovenia
e-mail: viktor.lovrencic@c-g.si

B. Al-Najjar
School of Engineering, Linnaeus University, Växjö, Sweden
e-mail: basim.al-najjar@lnu.se

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_33

1 Introduction

Over the past decade, plant maintenance has evolved to be one of the most important areas in the business environment [1]. The potential impact of maintenance at the level of operations and logistics (flexibility, throughput time, quality, etc.) is considerable, and therefore the financial implications of maintenance can be substantial [2]. Thus, even though in the past maintenance was usually considered just as a cost, nowadays it represents an important source of a competitive advantage [3]. The importance of effective maintenance is essential, given the fact that the costs of maintenance, estimated to be between 15 and 40 per cent of production costs [4]. On the other hand, in a study [5] author found that the mean percentage of OEE across the sampled cases was 55 percent. In this regard a study [6] indicated that the company can increase its production capacity without investing in new machinery if it implements an efficient maintenance policy, which allows enhancing availability strongly, quality rate and performance efficiency moderately. Maintenance is therefore a factor that affects OEE to a high extent and small investments in maintenance can lead to appreciable reduction in production cost [3, 7].

Using vibration-based Maintenance (VBM) it is possible to receive indications of changes of the condition of a machine at an early stage [8]. In addition, effective usage of vibration-based Maintenance (VBM) within Total Quality Maintenance (TQMain) concept for rotating machines provides the user with a long lead time for start planning cost-effective maintenance actions to avoid failures [9]. VBM is becoming more widespread, especially where downtime costs are high [6]. One of the aims of this chapter is to present VBM in the context of preventing mechanical failures as well as considering the potential benefits for company. Drawing on prior studies e.g. [3] that have examined the impact of VBM on company's business, our study aims to discuss the potential improvements of using VBM in the case of a paper mill company.

Recently, studies are beginning to address the live working in the electro-energetic sector [10]. There are great possibilities for the application of the live working (LW), especially from the perspective of achieving higher quality standards of the distribution of the electrical energy [10]. Consistently with the standard [11] this chapter adopts the following definition of the live working: all work in which a worker deliberately makes contact with live parts or reaches into the live working zone with either parts of his or her body or with tools, equipment or devices being handled. Hence, the idea behind the live working is discussed in this chapter in relation to paper machine, particularly in the context of achieving high availability and productivity as well as reducing the losses caused by unplanned stoppages. In this regard, the possibility of the outage-free maintenance brings an opportunity to contribute to the cost-effective maintenance.

This study therefore, examines the impact of mechanical and electrical failures on company's business on an example of a paper mill, using VBM and live working. The chapter is organized as follows. In Sect. 2, the theoretical background is provided. The case study analysis and results are presented in Sect. 3. Section 4 is devoted to discussion and conclusion.

2 Theoretical Background

The scope of maintenance in a manufacturing environment is illustrated by its various definitions. The British Standards Institute defines maintenance as [12]:

> A combination of all technical and associated administrative activities required to keep equipment, installations and other physical assets in the desired operating condition or restore them to this condition.

While maintenance is still often considered as a cost centre, research has highlighted the impact of the maintenance function on the company business [3,13]. In general, improvements in the performance of a maintenance policy aim to reduce production cost and increase company's profit and competitiveness through enhancing process availability, performance efficiency and quality rate [3]. The contribution of maintenance to the performance and profitability of manufacturing systems is to ensure that the plant can perform according to the agreed condition or what the company expected, by balancing between the allocation of maintenance resources and the plant output [14, 15]. Al-Najjar [16] stated that by applying effective maintenance, such as TQMain, a company's profitability and competitiveness can be enhanced through the continuous cost-effective improvement of production and maintenance processes. In this chapter we attempt to introduce a live working technique for reducing electrical failures, as well as to discuss a possibility to improve effectiveness of maintenance by using VBM (Fig. 1).

Figure 1 shows the basic interdependence between the maintenance and the production function. Essentially, by utilizing manufacturing system degradation occurs, and without proper maintenance this would lead to loss of function. Thus, the interaction between production and maintenance is very important. Negligence of maintenance and its role in the production process allows rapid degradation of machine and product quality [3].

On contrary to the VBM which is widely recognized maintenance approach, live working is still expanding in the field of maintenance of production systems. As we mentioned above live working method is any work in which the worker or his body parts, tools and equipment enter into contact with elements under voltage. In general, there are three methods of live working which help workers avoid the considerable hazards of live working [11]:

- *Hot stick working*: Hot sticks are used in live work by having the worker remain at a specified distance from the live parts and carry out the work by means of an insulating stick.
- *Working with insulating gloves*: In this type, the worker is electrically protected by insulating gloves and other insulating equipment, and carries out the work in direct mechanical contact with the live parts
- *Bare hands working*: This approach involves placing the worker in direct electric contact with the live parts.

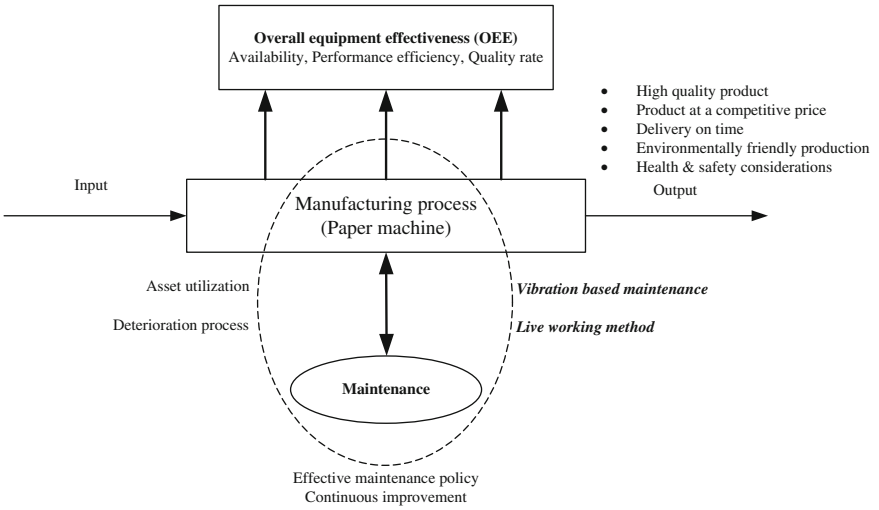


Fig. 1 The role of maintenance in cost-effective production

In electrical engineering, live working method is used for the maintenance of electrical equipment in order to prevent electrical failures. However, maintenance decisions regarding live working should be based according to the actual condition of the component (for instance, using infrared thermography to determine the condition of the component). Therefore, live working aims to ensure the reliability of electrical installations in production process by secure and continuous supply of electricity. Hence, continuous supply of electrical energy, smooth production processes (e.g. paper mill production processes), are identified as internal motivational factors of implementing live working method [17]. Apart from these definitions it is also necessary to indicate that safety requirements and other regulations must be considered when performing maintenance work under voltage. In addition to the legislation requirements, training of workers in the field of the live working is essential element in implementing the live working method [18].

2.1 Maintenance Performance Measures

In order to ensure a good performance of the production plant, organization needs to follow up the performance of maintenance processes [19]. It is essential to establish a proper performance measurement system in order to sustain improvement processes in companies [20]. Different measures are used in the field of maintenance. For the purpose of this study two measures are proposed. In the following a measure used to display the technical and economic impact of a more efficient maintenance policy on manufacturing process effectiveness is presented.

This equation presents the ratio between total maintenance-related cost and total accepted product [21]:

$$\frac{\textbf{total maintenance related cost}}{\textbf{total accepted product}}. \quad (1)$$

In this study total maintenance-related cost consist of direct labour, outsourcing and material. The second proposed measure represents the ration between maintenance cost and operating time:

$$\frac{\textbf{total maintenance related cost}}{\textbf{operating time}}. \quad (2)$$

This is to achieve a measure of maintenance cost that can be used independent of the machine operating time, because longer (or faster) use of a manufacturing machine usually increases maintenance cost [21].

3 Case Study

The case study was conducted in a Slovenian paper mill company. This research project began in autumn 2012 and aimed to increase the knowledge of how to improve maintenance policy and to introduce the possibility of implementing a live working method. The selection of suitable case was primarily based on what could be learnt in relation to the impact of maintenance on company's business from the perspective of the VBM and the live working method. The case study was conducted at the paper machine PM5. It could be argued that maintenance is highly crucial for this company, since the paper machine is running 24/7.

Different data were gathered for the purpose of this study. It consisted of two parts—technical and economic. In this regard, several plant visits were conducted in order to obtain all the relevant data, and to discuss with maintenance manager. As the economic data were confidential, the data used in the analysis were changed.

3.1 Data Analysis and Results

Prior to empirical analysis the total stoppage time was divided according to the following categories: short stoppages, unplanned stoppages and planned stoppages (Fig. 2).

As shown in Fig. 2 the short stoppage constitutes the largest portion of the stoppage time in all three presented years. The planned stoppages were planned, in general, every sixth week to perform certain operational tasks. However, one should note that planned stoppage time doesn't include time spent for an annual overhaul.

The results revealed that about 1.8 % in 2011, 2 % in 2010 and 2.8 % in 2009 of the planned operating time, the machine was stopped due to the unplanned stoppages. The causes of the unplanned stoppages are illustrated in Fig. 3.

One can also see (Fig. 3) that unplanned stoppages in the case of mechanical causes are decreasing since 2009 onwards. The latter implies that company managed to avoid certain amount of failures, especially those that can be monitored by vibration signals.

The maintenance department of the company has been implementing the vibration monitoring for several years. In this company the vibration analysis is outsourced and the measurements, analysis and evaluation of the condition of equipment monitored are delivered by a sub-supplier. However, our analysis shows that effectiveness of VBM policy could be improved. Periodic measurements are usually performed once per month, and based on these measurements company receive a paper report containing the spectrum analysis of past and current measurements. Mapping the condition of equipment and assessing its damage severity depends on several variables (e.g. past data, trend, current measurement, deterioration rate, maximum allowable limit for the condition monitoring (CM) value, residual time and probability of failure at that CM value level) [7]. Having this in mind, it can be argued that maintenance decisions cannot be effectively supported based on these paper reports. Therefore, with a more accurate VBM recommendations company could plan replacements of bearings during the planned time, when a paper machine is not operating due to certain operational tasks conducting by the production department. This would lead to a reduction of unplanned stoppage time, an increase in availability, and enhancing the economic benefits that could be gained by more efficient maintenance. To support our discussion the detailed analysis of the year 2011 is presented in the following. Using the Eq. 1 we found that there are no major differences in the presented years. The costs slightly decrease from 2009 onwards. However, in year 2011 one can find somewhat

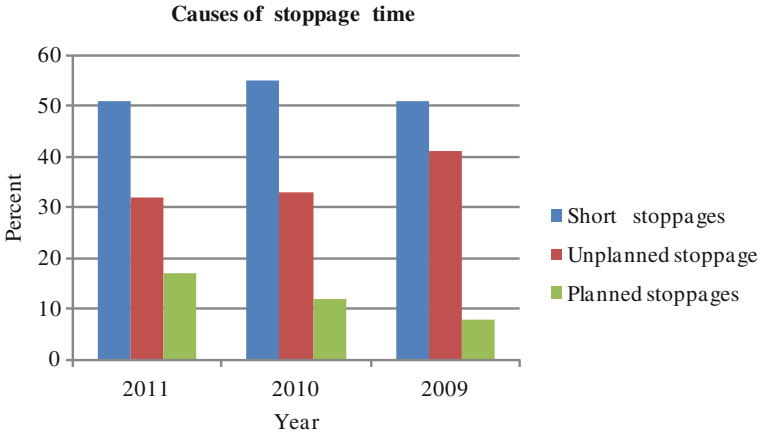


Fig. 2 Causes of total stoppage time at paper machine PM5

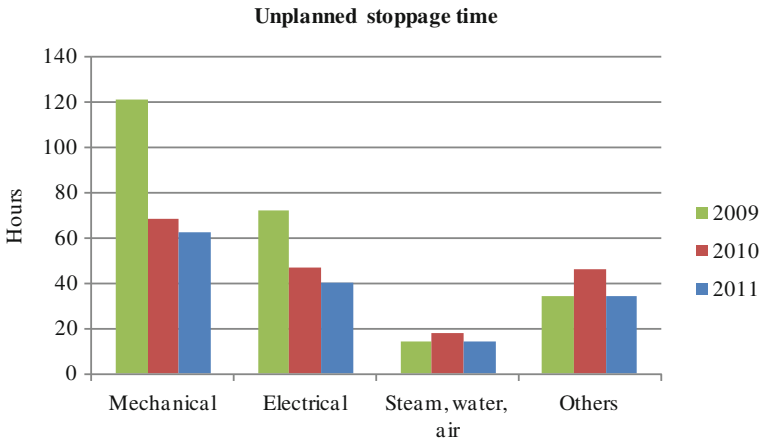


Fig. 3 Causes of unplanned stoppages

reduced productivity, mainly due to low demand as will be explained later. Using Eq. 1 we assessed the maintenance cost with respect to operating time (Fig. 4).

As shown in Fig. 4, the highest value corresponds to December, November, September, followed by February and January. The plausible explanation for the highest value is that during this month conducted an annual overhaul, which ultimately resulted in higher costs and lower operating time. Further, the second highest value was caused by the low demand, which decreased the operating time. The results revealed that the majority of unplanned stoppages occurred during the January. As such this period was taken into further analysis. It was found that mechanical causes prevailed alongside all unplanned stoppages. In particular, two

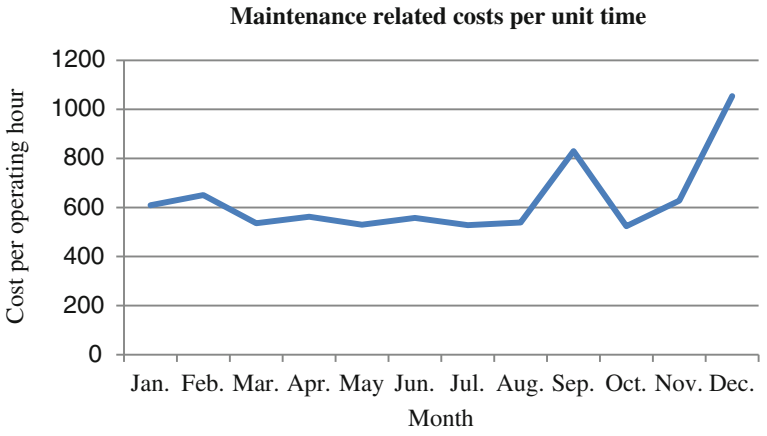


Fig. 4 Maintenance related costs per unit time in year 2011

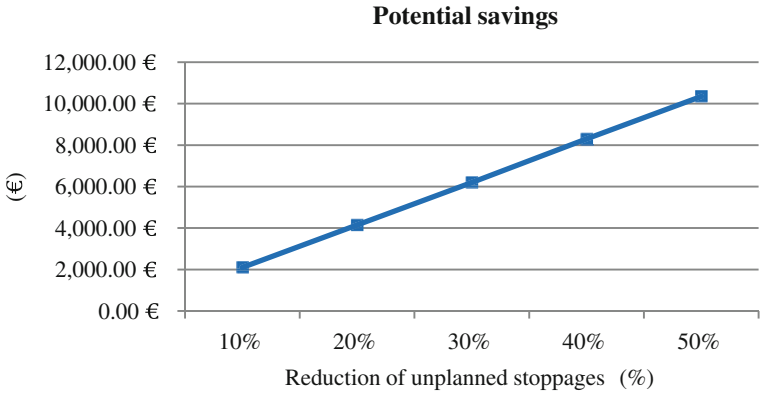


Fig. 5 Potential savings of unplanned stoppages (electrical failures) reduction in year 2011

unplanned stoppages could be avoided using vibration monitoring. These failures could be prevented by using more effective VBM. This means that mechanical components could be replaced either in November during the time planned for regular operational tasks or in the end of the December while an annual overhaul was performed. In addition we assessed the financial impact of maintenance in the case if more effective maintenance policy would be implemented. Hence, expressed in financial value this means 16.000,00 € profit losses (potential savings).

Second part of this case study includes an analysis and discussion of possible benefits of live working method. This method was introduced to the company and some minor tasks were already performed. With this method maintenance tasks can be conducted without having previously disconnected the energy. It is mainly aimed to reduce stoppages due to electrical causes. For the purpose of this study we performed the simulation (for the year 2011) of the reduction of unplanned stoppages, as illustrated in Fig. 5.

As shown in Fig. 5, we presented profit losses (potential savings) in the case if unplanned stoppages due to electrical causes would be reduced from 10 to 50 %.

4 Discussion and Conclusion

The primary purpose of this study was to illustrates/discusses the potential savings of mechanical and electrical failures due to more efficient maintenance policy. Results of this study are consistent with the studies [3, 6, 13] in which authors presented how an effective maintenance policy could influence the company's business. Thus, taking into account the fact that manufacturing performance can be influenced by maintenance policy, this study therefore implies that if company would be more successful in equipment condition assessment, more failures could be prevented. In this chapter we demonstrated of how more efficient VBM and the

deployment of live working could impact the company's business considering the profit losses/potential savings. Using the technical and economic data together allowed us to achieve a better overview on the maintenance from the perspective of cost-effective production. However, to assess the cost-effectiveness of a manufacturing process, the savings could be compared to the investment. We did not examine all mechanical failures from year 2009 to year 2011. Nonetheless, our results show that in the presented month (January 2011) investment in VBM could be cost-effective, if mechanical failures would be prevented as illustrated above. Our findings are consistent with study [22], in which author explores the importance of condition-based maintenance in achieving company's profitability and competitiveness. It is clearly evidenced that company could benefit from using more efficient maintenance policy. For instance, the losses expressed in profit were 16.000,00 €, considering only in the observed period (January 2011).

Besides the potential benefits of using more efficient VBM, our study revealed that company could also benefit from implementing the live working method. The simulation of the potential savings showed that savings range from 2.000,00 to 12.000,00 €, depending on level of the reduction of unplanned stoppages caused by electrical failures in year 2011. Our study underscores previous assertions that live working provides an important attribute of achieving an efficient production process [18]. In this sense, our study contributes to the literature by providing new empirical evidence of the potential savings in the context of live working method. Considering the complexity of this method, our study indirectly support the findings of Crespo Márquez et al. [23], who suggest that higher levels of knowledge, experience and training is required, and at the same time, techniques covering the involvement of operators in performing maintenance tasks are extremely important to reach higher levels of maintenance quality and overall equipment effectiveness.

We summarize the contribution of the chapter as follows:

- If properly implemented, VBM has the potential to reduce the unplanned stoppages caused by mechanical failures. Our findings further support the proposition that an effective VBM positively impacts the company's business.
- Live working is an effective maintenance method to increase the availability of the paper mill production process. The results indicated that reducing the unplanned stoppages caused by electrical failures, could also be associated with company's positive business performance.

Overall, we believe that potential savings are considered to be sufficient to trigger companies to adopt an effective maintenance policy. The other major management implication is that companies should be able to determine in which circumstances, if at all, the implementation of certain maintenance policy is needed. It is suggested that both technical as well as economic data are necessary for inducing a decision on the adoption of maintenance policy.

References

1. Kutucuoglu KY, Hamali J, Iran I, Sharp JM (2001) A framework for managing maintenance using performance measurements systems. Int J Oper Product Manage 21(1):173–195
2. Waeyenbergh G, Pintelon L (2002) A framework for maintenance concept development. Int J Prod Econ 77(3):299–313
3. Al-Najjar B (2007) The lack of maintenance and not maintenance which costs: a model to describe and quantify the impact of vibration-based maintenance on company's business. Int J Prod Econ 55(8):260–273
4. Dunn R (1987) Advanced maintenance technologies. Plant. Engineering 40:80–82
5. Ljungberg O (1998) Measurement of overall equipment effectiveness as a basis for TPM activities. Int J Oper Product Manage 18(5):495–507
6. Al-Najjar B, Alsyouf I (2004) Enhancing a company's profitability and competitiveness using integrated vibration-based maintenance: a case study. Eur J Oper Res 157(3):643–657
7. Al-Najjar B (2012) On establishing and running a condition-based maintenance policy; applied example of vibration-based maintenance. J Qual Maint Eng 18(4):401–416
8. Al-Najjar B (1997) Condition-based maintenance: selection and improvement of a cost-effective vibration-based policy in rolling element bearings. Doctoral thesis, ISSN 0280-722X, ISRN LUTMDN/TMIO—1006—SE, ISBN 91-628-2545-X, Lund University, Institution of Industrial Engineering, Sweden
9. Al-Najjar B, Ciganovic R (2009) A model for more accurate maintenance decisions (MMAMDEC). In: Kiritsis D, Emmanouilidis C, Koronios A, Mathew J (eds) Proceedings of the 4th world congress on engineering asset management (WCEAM). Athens, Greece 28–30 Sept, pp 7–14
10. Lovrenčič V, Lušin M (2008) Uvajanje dela pod napetostjo (DPN) v slovensko elektroenergetsko okolje [Introduction of the live work in the Slovenian electro-energetic environment], 8. Mednarodna konferenca Globalna varnost, Vodnik po konferenci, 13.-14.11.2008, Brdo pri Kranju
11. SIST EN 50110-1:2007 (2007) Operation of electrical installations, SIST, January 2007. Rules on industrial safety with regard to electric current hazards, Ur. l. RS, no. 29/1992
12. BSI (1984) Glossary of maintenance terms in terotechnology. British Standard Institution (BSI), London; BS 3811
13. Maletič D, Maletič M, Al-Najjar B, Gomišček B (2012) The role of maintenance regarding improving product quality and company's profitability: a case study. In 2nd IFAC workshop on advanced maintenance engineering, services and technology, Universidad de Sevilla. Sevilla, Spain. November 22–23, 2012, pp 7–12
14. Cholasuke C, Bhardwa R, Antony J (2004) The status of maintenance management in UK manufacturing organisations: results from a pilot survey. J Qual Maint Eng 10(1):5–15
15. Kelly A (1997) Maintenance strategy—business-centered maintenance. Butterworth-Heinemann, Oxford
16. Al-Najjar B (1996) Total quality maintenance. An approach for continuous reduction in costs of quality products. J Qual Maint Eng 2(3):4–20
17. Lovrenčić V, Oman V (2012) Nadgradnja sistema managementa kakovosti ISO 9001:2008 z zahtevami za izvajanje dela pod napetostjo [Upgrading the quality management system ISO 9001:2008 requirements for carrying out the live work]. In: Ferjan M, Kljajić-Borštnar M, Marič M, Pucihar A (eds). Quality, innovation, future. International conference on organizational science development, 21–23 Mar 2012, Portorož, Slovenia
18. Lovrenčič V, Lušin M (2011) Slovenian training experience and methodologies for live work implementation. International conference on live maintenance ICOLIM 2011, Zagreb, Republic of Croatia, May 31–June 2 2011
19. Parida A, Chattopadhyay G (2007) Development of a multi-criteria hierarchical framework for maintenance performance measurement (MPM). J Qual Maint Eng 13(3):241–258

20. Jaca C, Viles E, Mateo R, Santos J (2012) Components of sustainable improvement systems: theory and practice. TQM J 24(2):142–154
21. Al-Najjar B, Hansson MA, Sunnegårdh P (2004) Benchmarking of maintenance performance: a case study in two manufacturers of furniture. IMA J Manage Math 15:253–270
22. Alsyouf I (2007) The role of maintenance in improving companies' productivity and profitability. Int J Prod Econ 105(1):70–78
23. Crespo Márquez A, Moreu de León P, Gómez Fernández JF, Parra Márquez C, López Campos M (2009) The maintenance management framework: a practical view to maintenance management. J Qual Maint Eng 15(2):167–178

A Joint Predictive Maintenance and Inventory Policy

Adriaan Van Horenbeek and Liliane Pintelon

Abstract New maintenance policies like condition monitoring and prognostics are developed to predict the remaining useful life (RUL) of components. However, decision making based on these predictions is still an underexplored area of maintenance management. Furthermore, maintenance relies on the availability of spare parts for replacement in order to reduce failure downtime and costs. Accurate predictions of component failure times can be used to improve both maintenance and inventory decisions. During the past decades, several joint maintenance and inventory optimization systems have been studied in literature. Compared to the separate optimization of both models, these publications reported a remarkable improvement on total cost due to joint optimization. However, the inclusion of RUL in joint maintenance and inventory models for multi-component systems has not been considered before. The objective of this chapter is to quantify the added value of predictive information (RUL) in joint maintenance and inventory decision making for multi-component systems considering different levels of inter-component dependence (i.e. economic, structural and stochastic). A dynamic predictive maintenance policy is developed, which optimizes both maintenance and inventory parameters while minimizing the long-term average maintenance and inventory cost per unit time.

1 Introduction

1.1 Literature Review

The joint optimization of the maintenance and inventory problem is regarded as a promising area for the development of maintenance optimization [14]. This is because interesting advantages can be obtained by this integrated view. Some

A. Van Horenbeek (✉) · L. Pintelon
KU Leuven, Centre for Industrial Management, Celestijnenlaan 300A, Louvain, Belgium
e-mail: adriaan.vanhorenbeek@cib.kuleuven.be

L. Pintelon
e-mail: liliane.pintelon@cib.kuleuven.be

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_34

authors advocate that when dealing with maintenance problems in a restricted way (i.e. not considering the spare parts availability) the results may be questionable since, in practice, the proposed policy cannot be adopted due to lack of spare parts in inventory. In fact, the availability of the spare-parts is one of the most important factors to avoid long downtimes of equipment [6]. For a thorough study of integrated policies of maintenance and inventory, see for example [8, 9, 14]. In this chapter we are specifically interested in joint models considering condition-based and predictive maintenance. Therefore, we present some particular, interesting works that consider these types of joint models below.

When measurements (i.e. condition monitoring) are used to estimate the condition of a component, the general life time distributions, which are based on an entire population of components, used in preventive maintenance models can be replaced by more realistic remaining lifetime distributions [7]. By dynamically updating the lifetime distributions after each inspection, more accurate information is available to set the replacement and spare ordering times. The authors [7] integrate the updated lifetime distributions into the model of [2] by considering a single-component system and single-unit storage capacity. To the best of our knowledge, this is the only chapter that incorporates RUL information into the joint maintenance and inventory decision problem. This might be striking because of the increasing importance of predictive maintenance in industry [14]. Furthermore, a reduction in spare parts and inventory cost is generally considered as one of the most important indirect benefits of a predictive maintenance strategy. Due to the available predictive information, component replacement can be anticipated and spare parts can be ordered "just-in-time".

1.2 Objective of the Chapter

We present a sequential optimization of both maintenance and inventory for a multi-component system with component interdependencies (i.e. economic and structural dependence [11]) taking into account predictive information (RUL). The predictive maintenance model presented in [15] is extended by the inclusion of an inventory policy as presented in [7]. In this way we want to provide insight in the joint maintenance and inventory problem for a multi-component system with component interdependencies considering predictive information on component degradation. We are specifically interested in the behavior of the proposed policy with regard to changing component interdependencies. This is because due to the component interdependencies and interactions maintenance actions will be grouped and the demand pattern for spare parts changes accordingly [15]. The joint policy optimizes both maintenance and inventory parameters while minimizing the long-term average maintenance and inventory cost per unit time. The added value of the joint predictive maintenance and inventory policy is compared, by means of a numerical example, to an age-based preventive maintenance policy without grouping joined with the same inventory policy as for the proposed predictive joint policy.

The structure of the Chapter is as follows. Section 2 describes the considered system and the degradation model. A brief overview of the predictive maintenance policy is given in Sect. 3, however more details on this policy can be found in [15]. The characteristics of the inventory policy are discussed in Sect. 4. Section 5 gives an overview of the considered component dependencies. Finally, a numerical example and conclusions are given in Sects. 6 and 7.

2 System and Degradation Model

Consider a series system with n non-identical components. A failure of component i causes the entire system to stop and a system and/or component failure is noticed immediately without any inspection. Maintenance is assumed to be perfect. Time is discretized with a sampling time τ. Component degradation information is retrieved at each inspection point $T_{insp,z} = z \cdot \varepsilon_i, z \in \mathrm{Z}^+$ and ε_i is defined as the inspection period for component i such that $\varepsilon_i = s\tau, s \in \mathrm{Z}^+$. In order to perform maintenance (i.e. assumed to be replacement) on one component of the system, the entire system has to be stopped, so that system downtime is accrued. Moreover, during this downtime due to maintenance, the deterioration of the non-replaced components remains unchanged. Details on the inventory policy are given in Sect. 4.

2.1 Degradation Model

The component degradation is characterized by a physical variable D_i with $i = \{1, ...,n\}$, where $\{D_i(t),\ t \geq 0\}$ is a stationary gamma process with shape parameter ν and scale parameter μ and the following properties [16]:

- $D_{\mathrm{i}}(0) = 0$
- $D_{\mathrm{i}}(\mathrm{t})$ has independent increments
- For $t > 0$ and $h > 0$, $D_{\mathrm{i}}(\mathrm{t} + \mathrm{h}) - D_{\mathrm{i}}(\mathrm{t})$ follows a gamma distribution with shape parameter ν and scale parameter μ

A component i is said to be failed when the degradation level D_i exceeds the failure threshold $D_{i,failure}$. This deterioration failure threshold $D_{i,failure}$ is, contrary to most of the used degradation models in literature, modeled as a random variable. This approach is believed to better model the real degradation process of components as the failure threshold $D_{i,failure}$ depends on the variable operating load, uncertain operating conditions and variable component strength. These factors make that each component fails at a variable degradation level $D_{i,failure}$, rather than when a fixed degradation threshold is reached. For each $t \geq 0$, the probability of failure in time interval $(0, t)$ can then be written as the convolution integral [1, 16]:

$$\begin{aligned}\Pr\{X(t) \geq Y\} &= \int_{x=0}^{\infty} f_{X(t)}(x)\Pr\{Y \leq x\}\, dx \\ &= \int_{x=0}^{\infty}\int_{y=0}^{x} f_{X(t)}(x) f_Y(y)\, dy\, dx\end{aligned} \tag{1}$$

where $X(t) = D_i(t)$ (i.e. the deterioration at time t, $t \geq 0$) and the probability density function of $D_i(t)$ is given by a gamma distribution with shape parameter ν and scale parameter μ [16], and $Y = D_{i,failure}$ has probability density function $fY(y)$. The random variable $D_{i,failure}$ is modeled by a Weibull probability distribution with shape parameter α and scale parameter β in analogy to [10, 12]. Based on the inspection of the current degradation level $D_i(t) = d_i^0$, the failure probability function $F_i(t)$ is computed by stochastic simulation of the degradation process over time. Each time new information on the current degradation level d_i^0 is available—e.g. by inspection—a prediction of the remaining useful life is made. This prognosis is used in the presented predictive maintenance policy as short-term information in order to schedule maintenance actions on a rolling-horizon.

3 Predictive Maintenance Policy

3.1 Need for Grouping Maintenance Activities

In order to take the economic and structural interdependencies between components in a multi-component system into account, grouping of maintenance actions should be considered to find an optimal maintenance policy. Therefore, the presented predictive maintenance policy [15] is based on a dynamic policy for grouping maintenance activities [17]. One specific preventive or corrective maintenance action can be performed on each component i of the system. A preventive maintenance action has a component-dependent cost $c_{i,p}$ and a system-dependent or set-up cost S. A corrective maintenance intervention has a component-dependent cost $c_{i,c}$ and a set-up cost S. The cost S is independent on the performed action and the number of actions at the same time (e.g. economic dependence). The component-dependent cost c_i depends on the preventive replacement time t and the time-to-failure $T_{i,F}$ of the considered component. The objective is to group maintenance activities to reduce the maintenance cost (total set-up cost). For each group G_j of n components a cost C_{Gj} is saved [15]:

$$C_{G_j} = (n-1) \times S - \sum_{i \in G_j} \left(c_i\left(t^*_{G_j}\right) - c_i\left(t_i^*\right)\right) \tag{2}$$

where $(n-1) \times S$ are the savings by grouping n maintenance actions and $c_i\left(t^*_{G_j}\right) - c_i\left(t_i^*\right)$ is the additional cost of shifting maintenance activity i from the

individual optimal time t_i^* to the optimal group maintenance time $t_{G_j}^*$. The predictive maintenance policy aims at finding the grouping structure that minimizes the maintenance cost on a finite planning horizon PH.

3.2 Predictive Information (RUL)

The proposed predictive maintenance policy is considered as a dynamic maintenance policy, as every time new information on the observed degradation $D_i(t)$ of a component becomes available, the prediction of the remaining useful life of the component is updated. The degradation model described in Sect. 1.2.1 is used to predict remaining useful life $F_i(t)\big|d_i^0$, by numerical evaluation of Eq. 1 for each component i, based on the current degradation d_i^0. The stochastic simulation procedure to determine $F_i(t)\big|d_i^0$ can be found in [15].

3.3 Individual Maintenance Optimization

First, an optimal maintenance date on an infinite horizon is determined by decomposing the multi-component maintenance problem into n single-component maintenance optimization models considering an age-based replacement policy. This decomposition approach allows the scheduling of many components [5]. An average use [17] of the components is assumed and the dependencies and interactions between the components are neglected at this stage. In this way, the savings from joint execution of maintenance activities are ignored. Both a short-term (t_i^*) and long-term ($t_{i,l}^*$) optimal maintenance time are determined. The short-term optimal maintenance time takes into account the current degradation d_i^0, while for the determination of the long-term optimal maintenance age no information on the degradation level is available.

For an age-based replacement policy the asymptotic cost, where t_i^* is the minimizing argument, is given by van der Duyn Schouten and Vanneste [13]. In our Chapter this is extended to include non-zero maintenance downtimes into the decision problem as follows:

$$\mathrm{C_i}(t|D_i(t)) = \frac{c_{i,p} + S + b_i\left(1 - \prod_{l=0}^{t-1} p_i^l\right)}{1 + \sum_{j=2}^{t}\prod_{l=0}^{j-2} p_i^l + \prod_{l=0}^{t-1} p_i^l \cdot t_{i,p} + \left(1 - \prod_{l=0}^{t-1} p_i^l\right)\cdot t_{i,c}} \tag{3}$$

where the empty sum equals zero and the empty product equals one. $c_{i,p}$ is the component dependent preventive maintenance cost and $b_i = c_{i,c} - c_{i,p}$, with $c_{i,c}$ the

component dependent corrective maintenance cost. t is the age at which preventive maintenance is performed, $t_{i,p}$ is the downtime due to a preventive maintenance action and $t_{i,c}$ is the downtime caused by a corrective action. p_i^l is the probability that component i survives the next period τ given that its age equals l at the beginning of the current period and $q_i^l = 1 - p_i^l$. For each individual component with current degradation d_i^0 and corresponding remaining useful life $F_i(t)|d_i^0$, an infinite-horizon age-based replacement policy is formulated to find the optimal maintenance time t_i^*. Each time new information becomes available, the maintenance schedule is updated.

3.4 Penalty Functions

Grouping maintenance actions results in shifting maintenance activities from their individual optimal maintenance time t_i^* or $t_{i,l}^*$, to the joint execution time $t_{G_j}^*$, which is defined as the optimal maintenance execution time of group G_j. There are two possibilities in shifting maintenance from their individual optimal times: for some components the failure probability will be increased by extending their useful life, while for others the useful life will be decreased. In order to define the effect of shifting maintenance actions from their optimal times, penalty functions are constructed. A penalty function h_i defines the expected additional cost of shifting the maintenance time from the optimal maintenance time t_i^* or $t_{i,l}^*$ for a component. Penalty functions for both the next optimal maintenance time t_i^*, based on the short-term information, as for the mth ($m > 1$) maintenance occurrence, based on the long-term optimal maintenance time $t_{i,l}^*$, are defined. The penalty function, by adopting a long-term shift [17] with Δt the shift from the optimal maintenance time and defined as $\Delta t = z\tau, \forall z \in \mathbb{Z}$, for the first maintenance action on component i is defined as [5]:

$$h_i\left(t_i^* + \Delta t\right) = \begin{cases} \sum_{j=t_i^*}^{t_i^*+\Delta t-1} \left(q_i^j b_i | d_i^0 - C_i^* | d_i^0\right) \prod_{l=t_i^*}^{j-1} p_i^l | d_i^0 \\ if \quad \Delta t \geq 0 \\ \sum_{j=t_i^*+\Delta t}^{t_i^*-1} \left(C_i^* | d_i^0 - q_i^j b_i | d_i^0\right) \prod_{l=t_i^*+\Delta t}^{j-1} p_i^l | d_i^0 \\ if \quad \Delta t \leq 0 \end{cases} \tag{4}$$

According to the long-term shift rule the execution interval of the first maintenance action is changed according to the predictive short-term information, while all future maintenance intervals remain $t_{i,l}^*$, the long-term optimal maintenance time. The penalty function for the mth maintenance action, with $m > 1$, on component i becomes:

$$h_i^m\left(t_{i,l}^* + \Delta t\right) = \begin{cases} +\infty, & \forall \leq e \\ \sum_{j=t_{i,l}^*+}^{t_{i,l}^*-1} \left(C_{i,l}^* - q_i^j b_i \middle| d_i^M\right) \prod_{l=t_{i,l}^*+\Delta t}^{j-1} p_i^l \middle| d_i^M, & \forall e < \Delta t < 0 \\ 0, & \Delta t = 0 \\ \sum_{j=t_{i,l}^*}^{t_{i,l}^*+\Delta t-1} \left(q_i^j b_i \middle| d_i^M - C_{i,l}^*\right) \prod_{l=t_{i,l}^*}^{j-1} p_i^l \middle| d_i^M, & \forall \Delta t \geq 1 \end{cases} \tag{5}$$

with e the floor of $((m-1).t_{i,l}^*/\tau)$ [4]. When a component i fails, the penalty function h_i of the failed component is defined as:

$$h_i\left(t_{i,F}\right) = \begin{cases} 0, & \forall t = 0 \\ +\infty, & \forall t > 0 \end{cases} \tag{6}$$

This means, when a component i fails, preventive maintenance actions on the other components can be performed during the downtime due to the failure of component i. Due to this assumption, opportunistic maintenance is thus included in the model.

3.5 Maintenance Activity Grouping

The aim is to group the maintenance activities on the planning horizon PH in order to minimize the maintenance cost on this planning horizon. The finite planning horizon is defined as:

$$PH = \max_{i \in (1,\ldots,n)} \left(\left(t_i^* + t_{i,l}^* + t_{i,p}\right), \varepsilon_i\right) \tag{7}$$

The parameter ε_i is defined as the prognostic horizon for component i. The prognostic horizon is the time between two consecutive predictions of remaining useful life of component i, based on newly available component degradation information.

Grouping of maintenance activities on PH can be done by using the defined penalty functions in Eqs. 4–6. Define $H_{Gj}\left(t_{Gj}\right)$ the group penalty cost function of group G_j when maintenance activities on components $i \in G_j$ are all performed at time t_{Gj} instead of their individual optimal times t_i^*. The savings Q_{Gj} by grouping maintenance operations $i \in G_j$ and executing them at time t_{Gj^*} can be calculated as:

$$Q_{G_j}\left(t_{G_j}^*\right) = \left(\left|G_j\right| - 1\right) \times S - H_{G_j}^* \tag{8}$$

If the savings Q_{Gj} are positive, the group G_j is cost effective, which means it is better to group the maintenance actions rather than performing them at their optimal

individual times t_i^*. The final objective is to find the grouping structure GS_k that minimizes the total maintenance cost on the planning horizon PH. An adapted version of the grouping algorithm developed by Wildeman et al. [17] is used heuristically to find the optimal grouping structure GS_k [15].

3.6 Maintenance Execution and Rolling Horizon

Based on the previous step a maintenance schedule on the planning horizon PH is constructed. Maintenance actions are executed according to the maintenance schedule. A rolling-horizon approach is considered as each time the planning horizon is shifted and the maintenance schedule is updated by including newly available information on component degradation and the corresponding remaining useful life.

4 Inventory Policy

A system with n non-identical components and consequently non-identical spare parts is considered. For each component at most one spare component can be in stock or on order at any time. In this way, we extend the model of Elwany and Gebraeel [7] by considering a multi-component system with interdependencies. However, we adopt the same approach for the inventory policy, which is defined by replacing the traditional failure time distributions in the inventory model of Armstrong and Atkins [2] by the predictive information (RUL). At each updating time t_i^0 the remaining useful life $F_i(t)\big|d_i^0$ of component i is updated. This is used to determine an optimal maintenance schedule as described in Sect. 3. The optimal spare ordering time can be determined at each updating time t_i^0 based on the optimal replacement time t_i^* according to the sequential approach proposed by Armstrong and Atkins [2] by defining the Joint Cost Function (JCF) as follows:

$$JCF(t_o) = \frac{c_{i,p} + S + b_i F_i^0(t) + c_s \int_{t_{o,i}}^{t_{o,i}+L_i} F_i^0(t)dt + c_h \int_{t_{o,i}+L_i}^{t_i^*} \bar{F}_i^0(t)dt + c_o}{t_{i,p}\bar{F}_i^0(t) + t_{i,c}F_i^0(t) + \int_0^{t_i^*} \bar{F}_i^0(t)dt + \int_{t_{o,i}}^{t_{o,i}+L_i} F_i^0(t)dt + t_i^0} \quad (9)$$

where $F_i^0(t)$ equals $F_i(t)\big|d_i^0$. A fixed lead time L_i is considered for each component i. Denote $t_{o,i}$ as the scheduled time to order a spare for component i, where $t_{o,i} + L_i \leq t_i^*$. If a component fails before t_i^* it is replaced immediately if a spare is available, or else as soon as a spare arrives. If the component fails before $t_{o,i}$ an order is placed immediately. If the system is down due to a lack of spare parts, a shortage cost c_s per unit time is incurred. A cost of c_h per unit time is incurred for holding one

spare part in stock for one unit time and at each order an ordering cost of c_o is incurred. Furthermore, from the moment on a spare part is ordered the timing of replacement remains fixed and the RUL of the component is not updated anymore.

5 Component Dependencies

In order to be able to determine the performance of the proposed predictive maintenance policy when considering different levels of dependence (e.g. partial dependence) between the components, a dependence parameter α_d is introduced. This parameter α_d reflects the advantage of performing maintenance on multiple components at once compared to maintenance on a single component, in other words it affects the set-up cost S by adapting the savings Q_{Gj} (see (8)) when grouping maintenance as follows:

$$Q_{G_j}\left(t^*_{G_j}\right) = \alpha_d \times \left(\left|G_j\right| - 1\right) \times S - H^*_{G_j} \tag{10}$$

The dependence parameter α_d is assumed to incorporate the effect of both economic and structural dependence between the components in the considered system. The dependence parameter α_d ranges from 0 (0 %) to 1 (100 %), where $\alpha_d = 0$ means no economic and/or structural dependence, $\alpha_d = 1$ means maximal economic and/or structural dependence between the components and $0 < \alpha_d < 1$ corresponds to partial dependence. The detailed formulation of the set-up cost S can be found in [15].

6 Numerical Example

To determine the performance of the presented joint predictive maintenance and inventory policy, it is compared to an age-based policy partnered with the same inventory policy as given in Sect. 1.4. Under this policy, a unit is always maintained at its age $T_{a,i}$ or failure, whichever occurs first, where $T_{a,i}$ is a constant [3]. The objective of all policies is to minimize the long-term mean cost per unit time, defined as C^*.

6.1 Input Data

Consider a three component system ($n = 3$) with n non-identical components. Time is discretized with a period τ equal to one and $\varepsilon_i = 5$. The component degradation parameters, as described in detail in Sect. 1.2.1, are given in Table 1. The corresponding cost and time parameters for all components are shown in Table 2. t_{wait}

Table 1 Component degradation parameters

Component n	ν_i	μ_i	α_i	β_i
1	2.00	1	100	20
2	0.40	0.2	100	3
3	0.32	0.2	100	3

Table 2 Cost and time parameters

$c_{i,p}$	$c_{i,c}$	t_{wait}		t_{repair}		t_{inst}		t_{secD}	
		μ	σ	μ	σ	μ	σ	μ	σ
605	5,805	6	0.5	2	0.5	2	0.5	5	0.5
665	5,865	6	0.5	2	0.5	2	0.5	5	0.5
475	5,675	6	0.5	2	0.5	2	0.5	5	0.5

stands for the waiting time, $t_{replace}$ for the actual replacement time, t_{inst} for the installation time and the start-up time of the system and finally t_{secD} stands for the time to repair secondary damage. All these parameters determine the downtime due to preventive maintenance $t_{i,p}$ ($t_{replace}$, t_{inst}) and corrective maintenance $t_{i,c}$ (t_{wait}, $t_{replace}$, t_{inst}, t_{secD}). The cost of working (70), cost of transportation (120), downtime cost rate (400), shortage cost c_s (400), holding cost c_h (150) and order cost c_o (100) are also considered in the numerical example and are defined on a per unit or unit time basis. L_i equals (1, 2, 3).

6.2 Results and Discussion

The long term mean maintenance, inventory and total cost for both considered maintenance policies and different levels of dependence (α_d) are shown in Fig. 1. When no dependence exists ($\alpha_d = 0$) the predictive policy leads to a decrease in both maintenance and inventory costs. The predictive information (RUL) allows one to better schedule maintenance based on the real degradation of the components and at the same time allows one to order spare parts "just-in-time". Moreover, due to the predictive information less maintenance activities (corrective and preventive) are performed (i.e. the component useful life is extended), which results in a lower demand for spare parts and this reduces inventory costs. When we introduce dependencies between the components ($\alpha_d > 0$), the reduction in total cost of the predictive policy compared to the age-based policy becomes bigger. This is because the predictive policy considers the component interdependencies and groups maintenance activities; while the age-based does not consider component interdependencies when planning maintenance (i.e. the age-based joint policy is independent on the dependence) [15]. When looking in detail to the results of Fig. 1, we

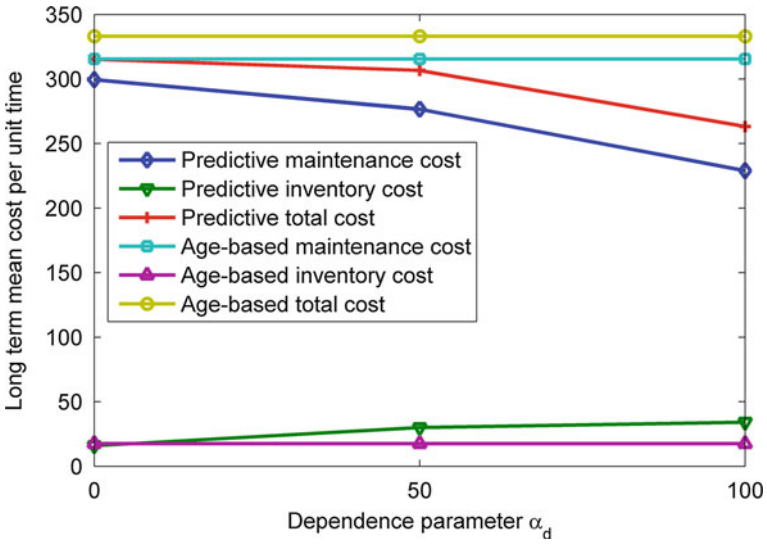


Fig. 1 Long term mean maintenance, inventory and total cost in relation to the dependence parameter α_d for both the joint predictive maintenance and inventory policy and the joint age-based maintenance and inventory policy

see that, as expected, the maintenance cost decreases as α_d increases, but on the other hand the inventory costs increase as α_d increases.

In other words, even with better predictability of the spare part demand, the inventory costs for the predictive policy are higher for a system with dependent components compared to the inventory costs for the age-based policy. The reason for this can be found in the changing demand for spare parts pattern due to the grouping of maintenance activities in the predictive policy when $\alpha_d > 0$. As maintenance grouping becomes more cost effective when α_d increases, more maintenance actions will be performed in a grouped way in order to save set-up costs. Also for the chosen degradation and cost parameters it is cheaper to shorten the component lifetimes instead of extending them to carry out a grouped replacement, which means that the demand for spare parts will rise as α_d increases. Our proposed joint predictive maintenance and inventory policy is a sequential policy where first the timing and grouping of maintenance actions are optimized and based on this the inventory decisions are optimized. As at the first stage of determining the optimal maintenance policy the inventory considerations are ignored, all advantages of the predictive information are reflected in the decrease in maintenance cost. In fact, the maintenance decision determines the inventory policy. However, the maintenance policy does not take into account the effects of an increased demand for spare parts on the inventory costs. This increased demand results in a burden on the inventory costs, as more orders need to be placed and the holding costs increase as we need more spares. The results clearly show that the use of predictive information in a joint maintenance and inventory policy has the

capability to reduce the costs significantly for multi-component systems with dependence. Although, the potential to reduce the costs of a joint predictive policy even further is present by optimizing both maintenance and inventory decisions jointly rather than sequentially (as proposed by Armstrong and Atkins [2]).

7 Conclusions

A joint dynamic predictive maintenance and inventory policy for multi-component systems considering different levels of dependence (i.e. economic and structural) is presented. The joint policy optimizes both maintenance and inventory parameters while minimizing the long-term average maintenance and inventory cost per unit time. The results show that the developed joint predictive maintenance and inventory policy reduces the long-term total (i.e. maintenance and inventory) costs. Although the total costs decrease when the dependence increases, due to the adopted sequential optimization approach the optimal maintenance schedule determines the inventory decisions, which leads to an increasing inventory cost when the dependence between the components increases. The results indicate that a real joint optimization, opposed to the sequential proposed in the Chapter; of the maintenance and inventory decisions has the potential to reduce the costs for dependent systems even further.

References

1. Abdel-Hameed M (1975) A gamma wear process. IEEE Trans Reliab 24:152–153
2. Armstrong MJ, Atkins DR (1996) Joint optimization of maintenance and inventory policies for a simple system. IIE Trans 28:415–424
3. Barlow RE, Proschan F (1964) Comparison of replacement policies, and renewal theory implications. Ann Math Stat 35:577–589
4. Bouvard K, Artus S, Bérenguer C, Cocquempot V (2011) Condition-based dynamic maintenance operations planning and grouping. Application to commercial heavy vehicles. Reliab Eng Syst Saf 96:601–610
5. Dekker R, Wildeman RE, van Egmond R (1996) Joint replacement in an operational planning phase. Eur J Oper Res 91:74–88
6. Díaz A, Fu MC (1997) Models for multi-echelon repairable item inventory systems with limited repair capacity. Eur J Oper Res 97:480–492
7. Elwany AH, Gebraeel NZ (2008) Sensor-driven prognostic models for equipment replacement and spare parts inventory. IIE Trans 40:629–639
8. Kabir ABM, Al-Olayan AS (1996) A stocking policy for spare part provisioning under age based preventive replacement. Eur J Oper Res 90:171–181
9. Kennedy WJ, Wayne Patterson J, Fredendall LD (2002) An overview of recent literature on spare parts inventories. Int J Prod Econ 76:201–215
10. Kong MB, Park KS (1997) Optimal replacement of an item subject to cumulative damage under periodic inspections. Microelectron Reliab 37:467–472

11. Nicolai RP, Dekker R (2007) Optimal maintenance of multi-component systems: A review. complex system maintenance handbook. Springer, London
12. Park KS (1988) Optimal wear-limit replacement with wear-dependent failures. IEEE Trans Reliab 37:293–294
13. Van Der Duyn Schouten FA, Vanneste SG (1990) Analysis and computation of (n, N)-strategies for maintenance of a two-component system. Eur J Oper Res 48:260–274
14. Van Horenbeek A, Buré J, Cattrysse D, Pintelon L, Vansteenwegen P (2012) Joint maintenance and inventory optimization systems: A review. Int J Prod Econ 143:499–508
15. Van Horenbeek A, Pintelon L (2013) A dynamic predictive maintenance policy for complex multi-component systems. Reliab Eng Syst Saf 120:39–50
16. Van Noortwijk JM (2009) A survey of the application of gamma processes in maintenance. Reliab Eng Syst Saf 94:2–21
17. Wildeman RE, Dekker R, Smit ACJM (1997) A dynamic policy for grouping maintenance activities. Eur J Oper Res 99:530–551

Proposal of a Quality Index Applied to Fault Detection Method in Electrical Valves

Leonardo Bisch Piccoli, Renato Ventura Bayan Henriques, Clayton Rocha, Eric Ericson Fabris and Carlos Pereira

Abstract In modern Intelligent Maintenance Systems, the machine or equipment robustness also depends on its capability to automatically generate reliability and safety reports. This paper describes an approach to autonomously identify if a faulty signal report has been correctly classified. The proposed approach builds on our previous experience in developing embedded intelligent maintenance systems and helps in avoiding the occurrence of "false positive" interpretations, which means, when the maintenance system indicates a possible fault that does not occur. This index would be useful for real-time monitoring and evaluation on fault detection systems, taking into account several degradation model characteristics. In order to validate the proposed methodology a test bench was developed in a lab reproducing some common faults and degradation processes that may occur in the field. The proposed approach makes use of a data acquisition equipment to store information from sensors to monitor specific physical variables from mechanical components such as gears. A test sequence is applied to the valve control actuators with the following steps: a few seconds of faulty free operational cycle sensor data (which means the opening and closing operations are executing without failure) are collected and then a faulty system behavior is emulated changing some mechanical actuator parts to faulty ones. In the faulty emulation case, a malfunction event must

L. Bisch Piccoli (✉) · R. Ventura Bayan Henriques · C. Rocha · E. Ericson Fabris · C. Pereira
GCAR - Control Automation and Robotics Group, UFRGS - Universidade Federal do Rio Grande do Sul, Av. Osvaldo Aranha, 103, Bom Fim, Porto Alegre, RS CEP:90031-190, Brazil
e-mail: leonardo@piccoli.eng.br

R. Ventura Bayan Henriques
e-mail: rventura@ece.ufrgs.br

C. Rocha
e-mail: oifui@ig.com.br

E. Ericson Fabris
e-mail: eric.fabris@ufrgs.br

C. Pereira
e-mail: cpereira@ece.ufrgs.br

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_35

be identified and reported by the fault detection system. The preliminary results indicate that the index is extremely useful especially when the degradation stage of a system is below, for example, catastrophic failure or a predefined level.

Keywords Fault detection · Electric actuators · Prediction

1 Introduction

In critical systems, the adoption of proactive maintenance strategies, as the monitoring, prognosis, and diagnosis combined in an embedded system, is of great importance [1]. An example is the use of pipelines for transportation of oil or gas over geographically dispersed regions, on which one interruption or bad may result in potential danger not only to the ecosystem but also to human lives with potential financial loss.

In general, the oil and gas industry model activities rely on reserve exploration, production, oil refining and distribution. In the refining stage, the hydrocarbons that form oil are separated giving a rise to distinct products (petroleum sub-products). The distribution refers to the derivatives transportation activities from crude oil up to the sales points. The maintenance procedures in tanks, ships, and other kinds of contention systems usually demand the coating and control of the inner walls of these containers to avoid leakage, corrosion, and fire. Currently, in Brazil, the biggest maintenance activities are conducted by Petrobras which includes corrective, preventive, and predictive maintenance actions.

Considering a worldwide increasing on demand for dynamic, automated systems become hostages of time, ie, many systems are designed to work autonomously, without human interference and fault detection. This cases with embedded systems through powerful mathematical capabilities along with major developments in softwares.

There is still a bottleneck that relies on human performance: the maintenance. We are still in a research and development phase of mathematical methods and models that try to recognize faults, and in some cases, predict the occurrence of anomalies in environment systems like electrical valves. These strategies could allow the maintenance technicians to a proactive act in advance to predict the faulty part or system, mainly considering the large universe of machinery and equipment parts available nowadays.

Considering the various types of existing maintenance strategies like predictive, preventive, or and corrective, any one of them relies on inspecting all equipment, check, clean, and perform some predefined procedure [2]. Due to the increase in complexity from actuators and demands for longer uptime operations, maintenance strategies relying on anticipating problems and acting directly on these forecasted failures to avoid downtimes is a goal to look for instead of a traditional reactive maintenance strategy.

They are critical in the oil and gas industry, while a bad manufacturing machine may produce many defective products. For example, this results in electrical valve failure and could cause production loss as regards intrinsic repaired cost. The investigation process measured signals or process models, for early detection of such possible machine failure before it really happens, is becoming compulsory.

The test-bench was designed and built to validate the proposed technique in real field applications failure using it as a test case in an electric valve. Therefore, a brake disk system was installed in a commercial fluid flow control valve system to simulate the open/close mechanical efforts. This approach makes it possible to monitor the torque or vibration effort delivered by the actuator through the gear mechanism during driving operations for opening and closing the valve, which could emulate typical initial failure situations on the gears before any catastrophic faults could occur.

It was used in the test setup a CSR6 model actuator from Coester company [3]. The valve actuator received torque and vibration sensors. Those sensors were installed at specific points permitting the collection of data at normal operating conditions and at failure as well. The torque signal was acquired for different operating conditions with variation on the dynamic behaviour of the system under test. The collected signals were processed by an embedded system with an algorithm implemented in ANSI-C language, employing digital signal processing techniques.

During valve operation there are several types of faults frequently happening from: gear faults, poor lubrication, hope valve stem, movable valve member break, rotor, and stator winding failures. The fault detection and diagnosis methods for electric for electric valve is based on self-organizing maps [4], correlating signals torque and vibration through sensor fusion and Fault Detection system. Since previous research have shown that more percent of faults in electrical valve are related to gear faults and movable valve member break which are common.

This paper describes a new condition of monitoring method for an electrical valve based on vibration and torque signal analysis. A robust bearing fault detection scheme has been developed by time domain feature extraction from vibration and torque signals of healthy and faulty gears.

The approach consists of two consecutive processes: fault detection process and fault diagnosis process as shown in Fig. 1. In the fault detection process, significant

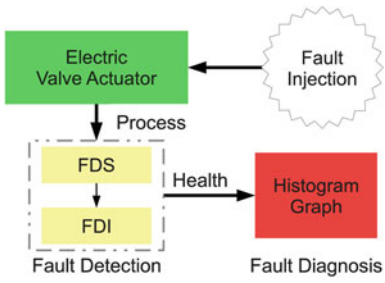


Fig. 1 Structure of model-based fault detection and diagnosis

features from vibration and torque signals are extracted through the Fault Detection System (FDS) algorithm to generate the fault state detection [5]. Consequently, this state is applied to the Fault Detection Index (FDI) to process and generate the pattern classification technique using histogram. However, instead of analyzing the vibration and torque signal to determine the valve faults, the signal can be classified to the corresponding faulty category in histogram.

The testing results show that our proposed approach provides significantly fault classification accuracy and a better performance than previous approaches [6] with assessment of actuator equipment condition [1].

2 The Proposed System

In this section, we will discuss the background topics necessary for understanding the proposed methods. An explanation of FDS, FDI, test bench and histogram mapping will be given, followed by some relationships between equations, transforms and graphs. Finally, our measure of enhancement will be introduced along with methodology for detection fault.

2.1 The Fault Detection System

One approach possible to detection fault is to use wavelet transform. The advantages of the wavelet transform over traditional transforms, such as the Fourier transform, are already very well known [7]. Due to its strait representation in time and frequency domains, it is widely used for linear signal analysis and pattern recognition. However, in environments with nonlinear and time-variant systems, the use of adaptive filters has become more attractive.

An LMS adaptive filter demands only multiply-add operations to be implemented which facilitates their usage in embedded systems. Moreover, the mathematical complexity is approximately the same for a FIR (Finite Impulse Response) filter. As FIR filters does not have feedback, the stability between input and output of a FIR filter is provided for any set of fixed coefficients and their implementation together with the LMS algorithm is relatively simple.

The FDS performed using the following signal processing steps: DTW, signal energy extraction and a LMS Adaptive Filter, as shown in Fig. 2. The DWT–LMS algorithm (Discrete Wavelet Transform Domain—Least Mean Square) has a better performance because it has an improvement in convergence speed and a reduction in steady state error [8]. Thus, this work proposes the development of an algorithm to be used in computers or micro-controllers based systems to perform fault detection and diagnosis system.

The DWT block is responsible for the original signal decomposition into several components located in time and frequency domain. Depending on the desired

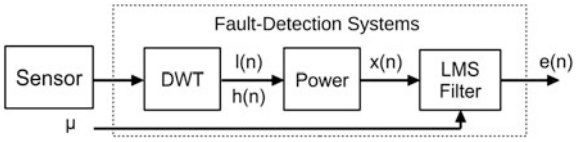


Fig. 2 Structure for the embedded system FDS

result, the coefficient of approximation *l(n)* or detail *h(n)* is sent to the block POWER, where the signal energy is extracted. Therefore, an adaptive filter LMS is used to predict values of the input signal *s(n)*, which are the desired response delayed by a specific number of samples. The adaptation parameter step-size μ controls the speed of convergence, stability, and steady state performance of the adaptive filter algorithm and also the sensitivity to obtain a transient response variation in the time domain system. It is important to be aware that in this application the output signal, (i.e. in other words the steady state error *e(n)*), represents the possible faults detection in the observed signal.

2.2 The Fault Detection Index (FDI)

This work presents an index that allows for evaluating monitoring and diagnosis performance of different fault detection signal from FDS, which takes into account from correct detection to non-detection during a fault event.

The FDI assigns ratio points each time the fault detection is signalized. The fault detection method evaluates the percentage ratio at every sample whether a fault in a FDS signal is occurring or not. The FDI as a sensitive fault detection index is analyzed into a feasible form, such data transmission over wireless networks like WirelessHART [9]. The FDI is defined by the equation:

$$FDI[x] = \frac{100}{N} \sum_{N*x}^{N*(x+1)-1} FD \tag{1}$$

In the equation for the arithmetic average percentage of fault detection FDI, *FD* is fault detection of each point in sample of FDS, *N* is the number of discrete measurements and *x* is sample of size *N*.

The FDI produces information about the mean and statistical fault detection recorded by the sensors acquisition subsystem. According to with this information stochastic storage and separation of these blocks, it is possible to transform an average of indexes on a reference of a part, component, or even a failure characterization of a code pre-analyzed.

The existence of a single index for evaluating fault detection performance that could allow screening different faults is of great importance for predictive aware maintenance systems. Overall, the proposed index plotted in a histogram is able to screen fault detection types.

2.3 Description of Test Setup

The equipment CompactRIO [10] NI cRI0-9004 of National Instruments together with the software LABVIEW [11] was employed in the Control, Automation and Robotics Laboratory (LASCAR) for data acquisition and control of the opening and closing processes of the electric valve actuator (Coester Automation model CSR6), as shown in Fig. 3.

The test setup was used for the preparation of test cases emulating common behaviour for the most common faults found in the real field application where the equipment becomes susceptible to the action of degradation such as aging, corrosion, cracks, damages caused by operators, etc.

It was used a disc brake in a test-bench for fault injection. This is installed on the actuator stem, and it was driven and regulated by means of a pneumatic valve controlled by an auxiliary actuator, as shown in Fig. 4. Thus, it was possible to perform an emulation of possible efforts that the valve suffers according to the passage of fluid in the pipe.

Figure 5 presents some possible gear aging conditions used to simulate several kinds of efforts and vibration delivered to the shaft depending on the gear health. Three different gears were employed, simulating field conditions for: standard gear, aged gear and fractured tooth gear.

Fig. 3 The developed test-bench

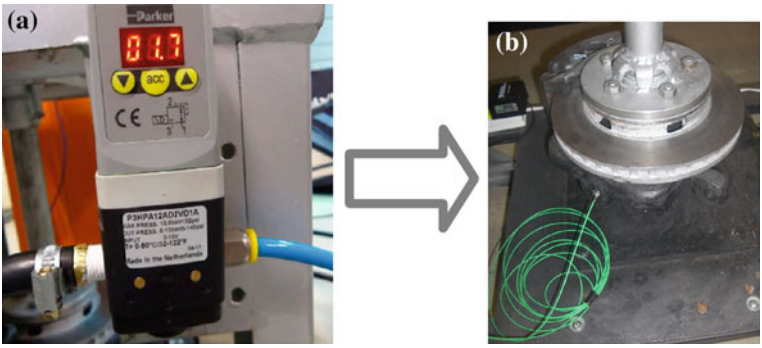


Fig. 4 Failures injection with disc brake

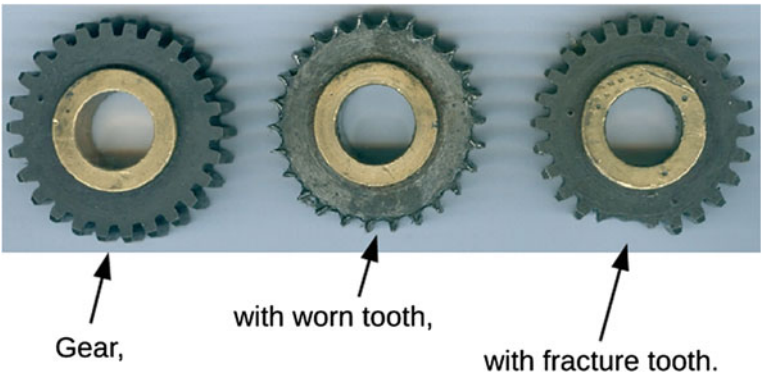


Fig. 5 Set of gears used in the tests

For those all the different operating conditions, sensors information was collected from the acquisition system using a vibration and torque sensor installed into the equipment as shown in Fig. 6, denominated as:

1. **Sensor 1**: Torque on the motor;
2. **Sensor 2**: Accelerometer on the motor.

In order to perform system tests, open/close test cycles of the valve were executed under normal and failure situations as it follows:

1. **Normal**: Normal test cycle and without external effort on the system;
2. **Fault type 1**: Test cycle with pressure applied through the brake set;
3. **Fault type 2**: Test cycle using two gears with aged tooth;
4. **Fault type 3**: Performed using a test cycle of three gears with fracture tooth.
5. **Fault type 4:** Performed using a test cycle with gears without lubrication.

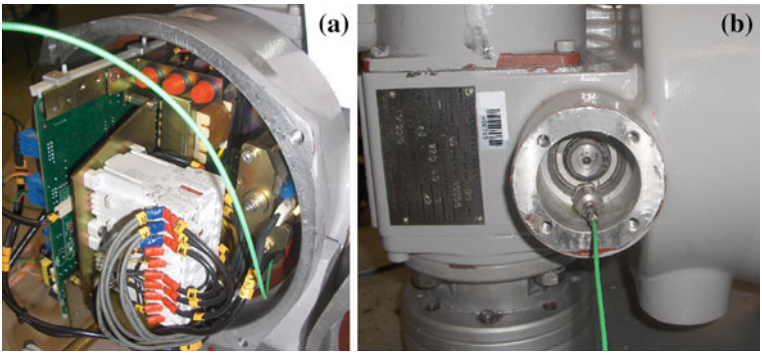


Fig. 6 Sensors on the engine compartment of the electric actuator

2.4 The Environment System

This work presents an index that allows to evaluate monitoring and diagnosis performance of different fault detection signal from FDS, which takes into account from correct detection to non-detection during a fault event. The current ripple and torque signal should be classified to the corresponding faulty category through a histogram.

Each of the fault injection process, fault detection process and fault diagnosis process are shown in Fig. 7 by the following steps:

1. **Step 1**: By artificially injecting faults of varying intensity into disc brake and swapping gears, we are able to study the machine operating in industrial plants work in heavy duty and long term degradation environments;
2. **Step 2**: The experiment consists of an electric valve actuator using a vibration and torque sensors installed into the equipment;
3. **Step 3**: Apply the SDF algorithm to those sensors to generate the fault detection signal. The algorithm detect the shift of the high sensor output values and frequency as an indication of the increase of fault;

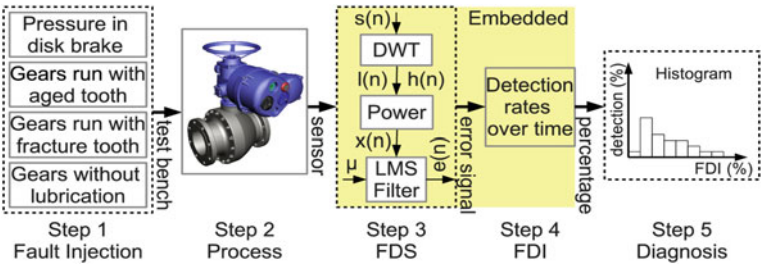


Fig. 7 Process flow of the proposed fault detection method

4. **Step 4**: The algorithm utilizes data obtained with the same sampling rate to construct high and low histograms of the SDF output signal. However, when a signal is sampled less than a Nyquist frequency there is a possibility that the sampling may occur at the transition between high and low signals or during only high or low signals.
5. **Step 5**: The detection performance of these methods, classify the signal to the equivalent faulty category based on the created histogram.

3 Results

Below, we discuss the detection performance of FDI for each type of fault. We describe how it detected faults in the corresponding situations. It was applied in three metrics to understand the performance of various faults: the faults detected, false negatives, and false positives.

In paper [5] it was shown that for certain time durations with medium and high impact intensity faults, there are no false negatives. For low intensity faults, a few false negatives could be reported at some times. Those false negatives for low fault intensity arise because the measurements with injected faults are very similar to the measurements without faults. On the other hand, when the valve operates under normal test condition and using the SDF method to detect faults, however in some worse cases show up a few false positives. The highest percentage number faults detected indicate the corresponding metric for the following tests.

For all test sequences, it was applied the following procedure: initially it was applied a few seconds of normal system cycle (no failure) followed by test cycles from specific failure. Thus, the data collected according the proposed fault detection method.

The histogram is used to produce visual information (frequency distribution) within FDI values for all processing tests. The histogram divides the series of FDI readings into 10 sample groups of percentages. Clearly, if the histogram has a certain mode, then the histogram based method will provide a fault diagnosis by selecting the type of fault.

3.1 Test for Fault Type 1

In this test it was used sensor 1 to feed the embedded system which executes the detection algorithm implemented in the ANSI-C language. Figure 8 shows the histogram generated by fault 1 with 1 bar and the second histogram show in Fig. 9 with 3 bars of pressure in disc the brake. In this test result the value used for parameter adaptation of the LMS adaptive filter was 0.1. In both cases, the energy was extracted from the approximation coefficients in FDS of the input signal.

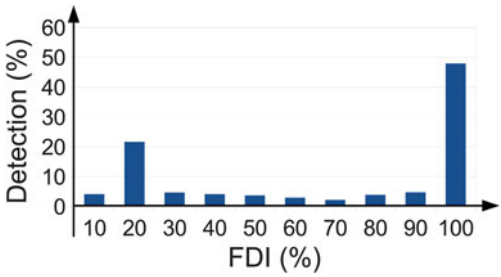


Fig. 8 Test with fault type 1, sensor 1, approximation coefficient, μ = 0.1 and 1 bar

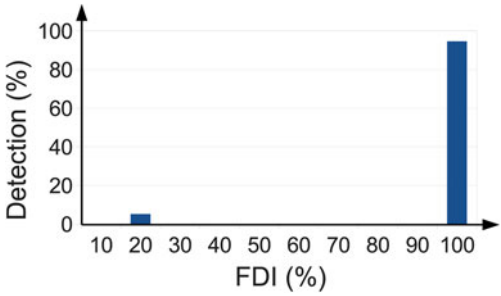


Fig. 9 Test with fault type 1, sensor 1, approximation coefficient, μ = 0.1 and 3 bar

It must be observed respectively the impact of medium and high duration fault intensity. The fault was detected for this type of failure in both opening and closing valve processes.

3.2 Test for Fault Type 2, 3 and 4

The following histogram use the same sensor 2 to generate the graphics. Similarly to the previous test, but now with the energy extracted from the detail coefficients signal. The error detected for each type of fault by the embedded system represented in histogram, corresponds a frequency of fault appearing in the actuator for each type of fault (Figs. 10, 11).

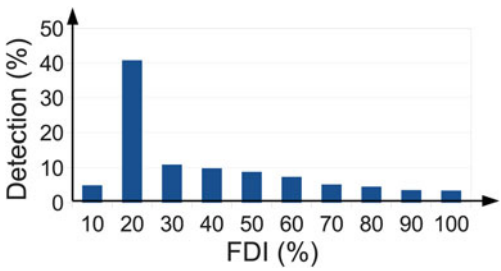


Fig. 10 Test with fault type 2, sensor 2, detail coefficient and μ = 0.1

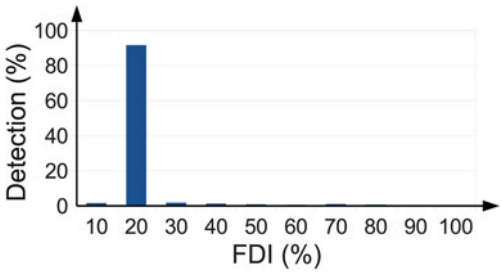


Fig. 11 Test with fault type 3, sensor 2, detail coefficient and μ = 0.1

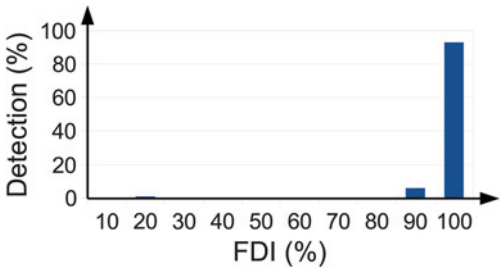


Fig. 12 Test with fault type 4, sensor 2, approximation coefficient and μ = 0.1

Similarly to the previous test, the same sensor 2 was used, the same coefficient of the LMS adaptive filter, but the energy was extracted by the approximation of the coefficient signal. The fault type 4 shows the duration fault intensity in Fig. 12.

4 Conclusion

This work presented as an embedded system description for fault detection using an an instrumented test-bench. A collection of tests were performed in the electric actuator coupled to a disc brake system, defective gears and poor lubrication for fault injection in order to replicate conditions under field. A database fault signature was created.

The histogram generated may be associated with a database correlating with a specific type of fault and this information could be sent to specific software as data input. This correlation between the FDS and the database parts could create an autonomous way to guide the technician as to which part needs to be replaced and at the same time send an order to the warehouse to tell which piece should be separated and sent to the maintenance sector.

It can not only predict fault but when compared with other business softwares it also ensures inventory acquisition, deliverance, and preparation of an identified faulty or aged part. That would mean greater integration between the purchasing department, which currently uses several procedures, but almost all with the use

of safety stocks or even zero inventory and maintenance departments that systematically require spare parts.

Therefore, in the efforts simulation the use of a test-bench together with the embedded system is of great importance to identify the many kinds of fault situations and degradation as well the appearance of cracks and component aging. Moreover, it will be possible to validate the system with the data collected on field and to embed all the electronics on a chip.

References

1. Lee J (1995) Machine performance monitoring and proactive maintenance in computer-integrated manufacturing: review and perspective. Int J Comput Integr Manuf 8(5):370–380
2. Pereira CE and Carro L (2007) Distributed real-time embedded systems: recent advances, future trends and their impact on manufacturing plant control. Annu Rev Control 31(1):81–92
3. Coester (2001) Coester Automation, Actuator Manual—Line CSR CSR 12- CSR 25 - CSR 50 Integral
4. Goncalves LF, Bosa JL, Balen TR, Lubaszewski MS, Schneider EL, Henriques RV (2011) Fault detection, diagnosis and prediction in electrical valves using self-organizing maps. J Electron Test 27(4):551–564
5. Piccoli LB, Henriques RVB, Fabris EE, Schneider EL, Pereira CE (2012) Embedded systems solutions for fault detection and prediction in electrical valves, in World congress on engineering asset management (WCEAM), out 2012
6. Lee J, Qiu H, Ni J, Ad Djurdjanovic D (2004) Infotronics Technologies and Predictive Tools for Next-Generation Maintenance Systems. 11th IFAC INCOM 2004. Salvador/Brazil
7. Murmu G, Nath R, Convergence performance comparison of transform domain lms adaptive filters for correlated signal, in devices and communications (ICDeCom), 2011 International Conference on, Feb 2011, pp 1–5
8. Konezny M, Rao S (1995) Improving the dwt-lms algorithm: boundary filter dwt matrix construction, signals, systems and computers. 1995 conferences Record of the twenty-ninth Asilomar conference on, vol 1, pp 75–81
9. Piccoli LB, Guimarães CSS Jr, Henriques RVB, Winter JM, Muller I, Netto JC, Pereira CE (2013) Embedded fault detection system using wirelessHART networks. NAVCOMP 1:1
10. Architectures RL, Practices D (2009) Compactrio developers guide, System
11. Sumathi S, Surekha P (2007) LabVIEW based advanced instrumentation systems, Springer Verlag

Selective Maintenance for Multi-state Systems Considering the Benefits of Repairing Multiple Components Simultaneously

Cuong D. Dao and Ming J. Zuo

Abstract Many industrial systems such as aircrafts, ships, manufacturing systems, etc. are required to perform several missions with finite breaks between missions. Maintenance is only available within the breaks. Due to the limitation of resources, all components in the system may not be maintained as desired. The selective maintenance problem helps the decision makers figure out what critical components to select and how to perform maintenance on these components. This paper studies the selective maintenance for multi-state series-parallel systems with the benefit of repairing multiple components simultaneously. Both time and cost savings can be acquired when several components are simultaneously repaired in a selective maintenance strategy. As the number of repaired components increases, the saved time and cost will also increase due to the share of setting up between components and another additional reduction amount from the repair of multiple identical components. A non-linear optimization model is developed to find the most reliable system subjected to time and cost constraints. Genetic algorithm is used to solve the optimization model. An illustrative example will be provided.

1 Introduction

Many systems in the industry are required to perform several missions with finite breaks between missions. For example, a manufacturing system works during weekdays and has a break on the weekend; an aircraft has hours to stop at the airport between consecutive flights; a ship has a few days' break before the next journey, etc. In general, we cannot do the maintenance on all components in the whole system. Therefore, we have to decide which components should be given maintenance activities within the limited resources between missions. This problem is called selective maintenance (SM). In the present work, the selective maintenance

C.D. Dao · M.J. Zuo (✉)
Reliability Research Lab, University of Alberta, 4–9, Mechanical Engineering Bld., Edmonton, AB T6G 2G8, Canada
e-mail: Ming.zuo@ualberta.ca

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_36

problem for multi-state series-parallel systems under strict requirements on resources such as time and cost is considered.

Since the late 1990s, selective maintenance has been attracting many researchers. The first model on SM was introduced in [1], in which binary-state series-parallel systems with independent and identical distributed (*i.i.d.*) components were studied. Then, different SM models for binary state systems were developed by Cassady et al. [2]. Schneider and Cassady [3] extended the model in [1] by considering selective maintenance for a fleet including multiple binary series-parallel systems. Maillart et al. [4] considered selective maintenance optimization for binary series-parallel systems working under multiple identical missions. Pandey et al. [5] studied the selective maintenance for a binary system under an age-based imperfect maintenance, in which the condition of components after maintenance depends on the cost spent and the maintenance may bring a component to a state which may be not "as good as new".

In [1–5], the system and its components are assumed to be in only two possible states of "failed" or "functioning". However, in practice, most systems and components can operate in more than two possible states. In this more general case, Chen et al. [6] provided selective maintenance model for multi-state systems with the objective of minimizing the cost of maintenance subject to reliability constraints. Liu and Huang [7] presented a selective maintenance model for a system with multi-state corresponding to cumulative performance of N binary components and considering the imperfect maintenance that may restore the condition of the system to an intermediate state. Pandey et al. [8] brought the concept of imperfect maintenance to a multi-state system with multi-state components and provided a selective maintenance model to maximize the reliability of the system in the next mission.

In all selective maintenance studies, [1–8], the repair of each component is considered to be independent from other components. However, in many industrial systems such as aircrafts, manufacturing systems, nuclear power plants, etc., repairing multiple components, especially *i.i.d.* components, is always more economical due to the share of setting up, tools, materials and labor. In this case, the repairs of those components in the system are economically dependent. Maaroufi et al. [9] considered maintenance for binary multi-component systems where a fixed "set-up cost" for dismantling and reassembling the system is incurred only one time when more than one components is replaced. Nourelfath and Chatelet [10], also studied the benefit of repairing multiple components of a parallel system in the production and preventive maintenance planning problem with the objective of minimizing the total production and maintenance cost. In these papers, they considered the set-up cost incurred each time of corrective, preventive or opportunistic replacement, where components are assumed to be "as good as new" after maintenance, rather than the possibility of repairing components to different intermediate states as in multi-state systems. Moreover, these papers did not consider time savings when performing maintenance on multiple components in the MSS.

In this paper, we will investigate the selective maintenance for multi-state series-parallel systems with multi-state components and the benefits of repairing multiple components simultaneously. Both time and cost saving can be assured when several components are selected to be repaired in a selective maintenance strategy. Two types

of economic dependency between repairing multi-state components based on the share of setting up and the advantage of repairing multiple *i.i.d.* components will be modeled.

2 Selective Maintenance Problem for the Multi-state Series Parallel System

The multi-state series-parallel system in this paper consists of M subsystems connected in series and there are N_i, $i = 1, 2, \ldots, M$ identical components connected in parallel in each sub-system (Fig. 1). Each component and the system may have $K + 1$ possible states from state 0 to state K, where state 0 is complete failure, state K is perfect functioning, others are intermediate states. The system performs consecutive missions with the break interval of T_0. The maintenance manager has to decide which components to maintain and how to maintain each component in the system within available budget C_0 and available time T_0.

We denote a component j in sub-system i by (i,j), $\quad j = 1, 2, \ldots, N_i$. For each component (i,j), $Y_{ij}, 0 \leq Y_{ij} \leq K$, represents its state at the time of entering the maintenance depot and X_{ij} represents its state at the beginning of the next mission. In this paper, we assume that the maintenance activities do not make the condition of components and the system worse, i.e. the states of the system and its components are not lower after exiting the maintenance depot. Thus, $Y_{ij} \leq X_{ij} \leq K$. In the selective maintenance problem, the state vector of all components in the system at the time of entering the maintenance depot, $Y = \{Y_{ij}\}$, $\quad i = 1, 2, \ldots, M$, $j = 1, 2, \ldots, N_i$, is known, we have to find its state vector at the time of exiting the maintenance depot, $X = \{X_{ij}\}$, $\quad i = 1, 2, \ldots, M$, $\quad j = 1, 2, \ldots, N_i$, that maximizes the system reliability in the next mission within available budget and time allotted between missions.

3 The System Reliability

The component degrades as the time of use and its state at the end of an operating mission is a random variable. The evaluation of components' performance degradation in multi-state systems has been investigated by Pandey et al. [8]. In this paper, we will not focus on the component's degradation process - readers may

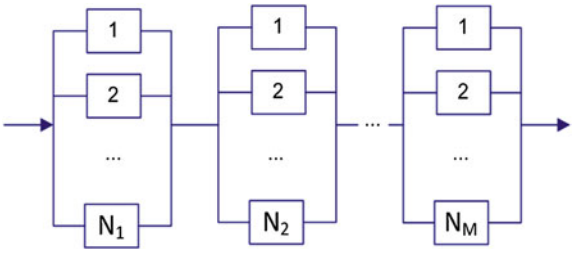


Fig. 1 Series-parallel system

refer to [8] for the state transition analysis of multi-state components and systems. It is assumed that the probabilities for a component in subsystem i at any pre-specified state b degrading to all possible state a ($0 \le a \le b$) after the operating mission, $p_i(b,a)$, are already known. When $b = 0, 1, \ldots, K$, these probabilities form a $(K+1) \times (K+1)$ transition probability matrix of each component in subsystem i for completing a mission as follows:

$$P_i = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ p_i(1,0) & p_i(1,1) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ p_i(K,0) & p_i(K,1) & \ldots & p_i(K,K) \end{bmatrix}, \quad i = 1, 2, \ldots, M$$

The summation of all elements on a row of the transition matrix is equal to 1. Thus:

$$\sum_{a=0}^{K} p_i(b,a) = \sum_{a=0}^{b} p_i(b,a) = 1 \; for \; b = 0, 1, \ldots, K \tag{1}$$

The system is required to work at a specified level b or above, i.e. the system' state ϕ_s is greater than or equal to b at the end of the operating mission. If the state of the system is less than the required working level, $\phi_s < b$, it will fail to complete the mission. We define the system reliability at state b, $R_s(b)$, as the probability $\Pr(\phi_s \ge b)$.

In series-parallel structure, the system is at a state greater than or equal to b when all subsystems are at state b or above. Furthermore, a subsystem is in a state less than b, $\phi_{subi} < b$, when all of its components are in states less than b. The event that the subsystem i to be in state b or above at the end of the next mission, $\phi_{subi} \ge b$, is the complement event of $\phi_{subi} < b$. Therefore, the reliability of the system at state b can be computed using the following equation.

$$\begin{aligned} R_s(b) &= \Pr(\phi_s \ge b) = \prod_{i=1}^{M} \Pr(\phi_{subi} \ge b) = \prod_{i=1}^{M} (1 - \Pr(\phi_{subi} < b)) \\ &= \prod_{i=1}^{M} \left(1 - \prod_{j=1}^{N_i} \sum_{a=0}^{b-1} p_i(X_{ij}, a) \right) \end{aligned} \tag{2}$$

4 The System Maintenance Time and Cost

4.1 Single Repair Time and Cost

Assume $t_i(a, b)$ and $c_i(a, b)$ are the required time and cost for a single repair of an *i.i.d.* component in subsystem i from state a to any state b, $b > a$. The single repair time and cost of a component i can be arranged in matrix form as:

$$T_i = \begin{bmatrix} 0 & t_i(0,1) & \ldots & t_i(0,K) \\ 0 & 0 & \ldots & t_i(1,K) \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & t_i(K-1,K) \\ 0 & 0 & \ldots & 0 \end{bmatrix}, \quad i = 1,2,\ldots,M$$

$$C_i = \begin{bmatrix} 0 & c_i(0,1) & \ldots & c_i(0,K) \\ 0 & 0 & \ldots & c_i(1,K) \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & c_i(K-1,K) \\ 0 & 0 & \ldots & 0 \end{bmatrix}, \quad i = 1,2,\ldots,M$$

In order to calculate the total repair time and cost of the entire system, it is necessary to analyze the relationship of the repair time and cost between components. When the repair time for each component is independent, the total maintenance time of the system is simply a summation of all the individual repair time of its components. If the state vector of components before maintenance, $Y = \{Y_{ij}\}, \quad i = 1,2,\ldots,M, \quad j = 1,2,\ldots,N_i$, is given, the total system maintenance time corresponding to a vector of component state at the time of exiting the maintenance depot, $X = \{X_{ij}\}$, can be represented as:

$$T = \sum_{i=1}^{M} \sum_{j=1}^{N_i} T_i(Y_{ij}, X_{ij}) \tag{3}$$

Similarly, the total system maintenance cost can be obtained by taking the sum of all single components' repair costs.

$$C = \sum_{i=1}^{M} \sum_{j=1}^{N_i} C_i(Y_{ij}, X_{ij}) \tag{4}$$

In (3) and (4), $T_i(Y_{ij}, X_{ij})$ and $C_i(Y_{ij}, X_{ij})$ are the corresponding elements in the time and cost matrix of components in subsystem *i*.

4.2 Time and Cost Savings When Repairing Multiple Components

Equation 3 and 4 show the total maintenance time and cost of the system when there is no advantage of repairing multiple components in a maintenance strategy. However, in most realistic systems, time and cost savings are achieved when multiple components are selected to be repaired, especially for identical components in the same subsystem. They usually require similar initial setting up and may

have the same process of repair, materials, tools and labor. In [9], a fixed "set-up cost" is incurred only one time when more than one components is replaced. In this paper, the concept of "set-up cost" is employed to both cost and time. In addition, it is usually more economical if multiple identical components in the same current state (condition) a are maintained to the same working state b, $b > a$. Additional time and cost savings for this type of repair should be addressed. In this section, we will focus on formulating the total actual repair time for a selective maintenance strategy based on the single repair time matrix with the consideration of two types of time savings aforementioned. Once the total repair time is found, the total cost of the system can be calculated accordingly.

In the first type of saving, a fixed amount of "set-up time" corresponding to an additional component, Δt_s (Fig. 2), is saved when many components are repaired in a selective maintenance strategy. For example, when repairing multiple components in a machine, we have to dismantle the machine only one time, do the same process of inspection to all components and reassemble the machine one time after completing all the maintenance activities. The more components to be maintained, the more time is saved due to the share of setting up. If N_r components in the system are maintained in a selective maintenance strategy, the total amount of time saved due to the share of setting up will be $(N_r - 1) \times \Delta t_s$.

Secondly, let's consider repairing multiple identical multi-state components in subsystem i with the individual repairing time of each component from state a to state b of $t_i(a, b)$. In addition to the time saved from setting up sharing, we introduce a time saving coefficient, $f_T^i(a, b)$, to represent the advantage of repairing multiple *i.i.d.* components in subsystem *i*. f_T^ican take any value between 0 and 1. When $f_T^i = 1$, there is no additional advantage of repairing identical components. $f_T^i = 0$ when we do not need any additional time to repair an extra component other than the first one. Let $t_i^{'}(a, b)$ be the adjusted repair time for an additional component in subsystem i from state a to state b. The adjusted repair time and the saved amount in addition to a fixed set-up time for repairing an *i.i.d.* component from state a to state b are calculated with (5) and (6) respectively.

$$t_i^{'}(a, b) = f_T^i(a, b) \times t_i(a, b) - \Delta t_s \tag{5}$$

$$t_i(a, b) - t_i^{'}(a, b) - \Delta t_s = \left(1 - f_T^i(a, b)\right) \times t_i(a, b) \tag{6}$$

The calculation of maintenance time considering the advantage of repairing multiple *i.i.d.* components is illustrated in Fig. 2.

The total system maintenance time is, then, the summation of total adjusted repair time of each component. It is a function of the decision variable X_{ij} if the single repairing time matrix, the time saved due to the share of setting up, the time dependent coefficients and the components' states at the time entering the maintenance depot are known.

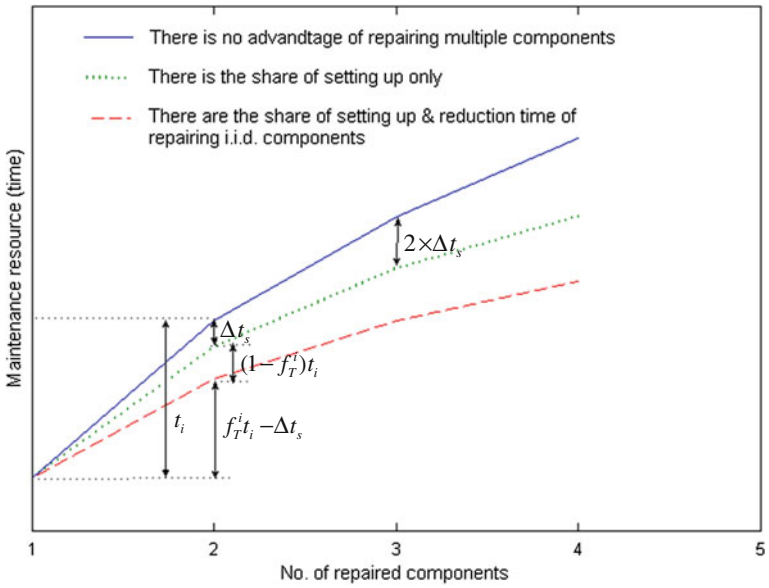


Fig. 2 The advantage of repairing *i.i.d.* components

$$T(X_{ij}) = \sum_{i=1}^{M} \sum_{j=1}^{N_i} T_i^{'}(Y_{ij}, X_{ij}) \tag{7}$$

Similarly, the total system maintenance cost for a selective maintenance strategy can be obtained if the amount of saved money due to the share of setting up—Δc_s, the opportunity cost coefficients—f_C^i and the components' state at the time of entering and exiting the maintenance depot are known.

$$C(X_{ij}) = \sum_{i=1}^{M} \sum_{j=1}^{N_i} C_i^{'}(Y_{ij}, X_{ij}) \tag{8}$$

5 The Selective Maintenance Optimization Model

In this paper, the maintenance manager has to find what maintenance activities associated with each component to be performed to achieve the maintenance objective of increasing the system reliability under limitation of resources such as time and cost. The selective maintenance problem can be formulated using a non-linear integer programming problem. The decision variables, X_{ij}, are the states of components at the time of exiting the maintenance depot. The system and components' characteristics such as the state of components at the time of entering the maintenance depot,

the time and cost of single repair, time and cost saving of multiple repair of each component, state probability distribution of each component in the next mission, etc. are explicitly known.

$$Maximize\, f = R_s(b) = \prod_{i=1}^{M}\left(1 - \prod_{j=1}^{N_i}\sum_{a=0}^{b-1} p_i(X_{ij}, a)\right) \tag{9}$$

$$Subject\ to: \qquad T(X_{ij}) \le T_0 \tag{10}$$

$$C(X_{ij}) \le C_0 \tag{11}$$

$$\begin{aligned} Y_{ij} \le X_{ij} \le K \\ X_{ij}\ is\ integer,\ i = 1, 2, \ldots, M;\ j = 1, 2, \ldots, N_i \end{aligned} \tag{12}$$

In the model, the objective function (9) is to maximize the reliability of the system at a specified working level b which has been formulated in Sect. 3. There are two types of constraints in (10) and (11) which restrict the total time and cost for all maintenance activities within available time, T_0, and budget, C_0. The repair time and cost of components may be dependent as explained in Sect. 4. Equation (12) sets the boundary on the components states at the time of exiting the maintenance depot. This equation implies that the maintenance does not worsen the state of a component, i.e. X_{ij} must be integer value between Y_{ij} and the maximum state K.

6 Solution Approach and Example

In this paper, we use genetic algorithm (GA) [11] to solve the proposed selective maintenance models. A solution of the selective maintenance problem is represented by a chromosome consisting of N genes, where N is the total number of components in the system, $N = \sum_{i=1}^{m} N_i$. Each gene represents the state of a corresponding component at the end of the maintenance break (Fig. 3).

All elements in the vector of components' state are ordered from subsystem *1* to subsystem M and the decision variable, $X_{ij}, \quad i = 1, 2, \ldots, N_i, \quad j = 1, 2, \ldots, M$, can be transformed to a state vector X with the dimension of $(1 \times N)$. If the states of components at the time of entering the maintenance depot are known and can be rewritten in vector form, $Y(1 \times N)$, we can use GA to find the best combination of the components' outcome states after the maintenance break for the proposed optimal selective maintenance problem.

Based on this solution approach, the problem is coded using Matlab R2012. An illustrative example and results are shown as follows.

Example Considering a multi-state series-parallel system consisting of two subsystems, 7 components as in Fig. 4. The multi-state system and its components can be in 5 different states from state 0 to M = 4.

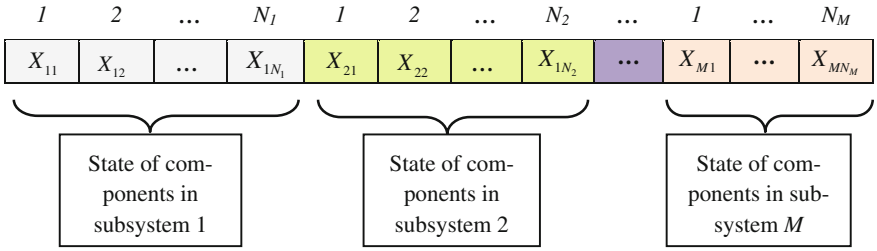


Fig. 3 Solution representation

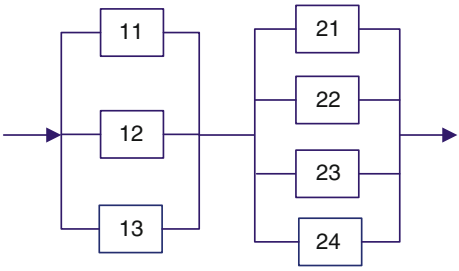


Fig. 4 MS series-parallel system

The transition probability matrices and corresponding cost and time matrices for each individual maintenance activity are:

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.15 & 0.85 & 0 & 0 & 0 \\ 0.15 & 0.1 & 0.75 & 0 & 0 \\ 0.05 & 0.15 & 0.15 & 0.65 & 0 \\ 0.05 & 0.05 & 0.1 & 0.1 & 0.7 \end{bmatrix}, P_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 & 0 \\ 0.1 & 0.25 & 0.65 & 0 & 0 \\ 0.05 & 0.1 & 0.1 & 0.75 & 0 \\ 0.05 & 0.1 & 0.1 & 0.2 & 0.55 \end{bmatrix}$$

$$C_1 = \begin{bmatrix} 0 & 2 & 4 & 6 & 8 \\ 0 & 0 & 3 & 5 & 7 \\ 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, C_2 = \begin{bmatrix} 0 & 3 & 4 & 7 & 10 \\ 0 & 0 & 2 & 5 & 8 \\ 0 & 0 & 0 & 4 & 7 \\ 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$T_1 = \begin{bmatrix} 0 & 2 & 5 & 7 & 8 \\ 0 & 0 & 3 & 4 & 6 \\ 0 & 0 & 0 & 3 & 5 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, T_2 = \begin{bmatrix} 0 & 2 & 3 & 6 & 9 \\ 0 & 0 & 2 & 4 & 7 \\ 0 & 0 & 0 & 3 & 6 \\ 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

All the available information of the system and its components such as the states of components at the end of the previous mission, time and cost saving due to the share of setting up and time and cost saving coefficients as well as the requirement of the system' working level and the resource available are presented in Table 1.

Using the input data in Table 1, we solve the SM model in Sect. 5 to find the optimal reliability of the system at state $b = 3$. The vector of component's state at the time of exiting the maintenance depot is $X = [4 \quad 0 \quad 4 \quad 3 \quad 3 \quad 1 \quad 3]$, that is, we select to repair components 1 and 3 in subsystem 1 to state 4, components (2, 1), (2, 2) and (2, 4) in subsystem 2 to state 3, do nothing to components (1, 2) and (2, 3). We also solve the SM model for the case that there is no benefit of repairing multiple components. Table 2 shows the results of the SM strategies on the multi-state systems.

It is clear from the results in Table 2 that more maintenance actions can be performed when considering the benefits of repairing multiple components simultaneously. When there is no benefit of repairing multiple components, we restore 4 components (1, 1), (1, 3), (2, 1), (2, 4) to an intermediate state 3 and do nothing to other 3 components (2, 2), (1, 2), (2, 3), while components (1, 1), (1, 3) can be maintained to the highest state (perfect functioning state) and another maintenance action can be performed on component (2, 2) (from state 2 to state 3) if the benefits of repairing multiple components are considered. Consequently, a significantly higher system reliability at level 3 can be achieved, $R_s(3) = 0.945$, and the resources utilization is also higher by 18.8 cost units and 14.6 time units when considering the advantages of repairing multiple components simultaneously.

Figure 5 illustrates the time (a) and cost (b) savings versus the number of components involved in a maintenance strategy. In this figure, we plot the resource usage for the selective maintenance strategy corresponding to vector $X = [4 \quad 0 \quad 4 \quad 3 \quad 3 \quad 1 \quad 3]$ in the example above and assume that components are maintained in an order from the first subsystem then the second subsystem.

In this maintenance strategy, 5 components including (1, 1), (1, 3), (2, 1), (2, 2), and (2, 4) are selected to be maintained simultaneously. When the benefits of repairing multiple components are considered, the total maintenance time and cost are smaller than those in the case of independent repairs. The amounts of time and cost savings increase considerably as the number of repaired component increases. The total time and cost required to repair all 5 components when considering the benefits of multiple repairs are just 14.6 time units and 18.8 cost units, while 21 time units and 26 cost units are required respectively when there is no advantage of repairing multiple components.

Table 1 Input data of the selective maintenance model

Y_{11}	Y_{12}	Y_{13}	Y_{21}	Y_{22}	Y_{23}	Y_{24}	Δc_s (cost units)	Δt_s (time units)	f_C^i $(i = 1, 2)$	f_T^i $(i = 1, 2)$	b	C_0 (cost units)	T_0 (time units)
1	0	1	2	2	1	2	0.3	0.4	0.6	0.6	3	20	15

Table 2 The results of the selective maintenance model

	Repairing action on component (i, j)				$R_s(3)$	$C(X_{ij})$ (cost units)	$T(X_{ij})$ (time units)
	(1, 1), (1, 3)	(2, 1), (2, 4)	(2, 2)	(1, 2), (2, 3)			
(*)	$1 \rightarrow 3$	$2 \rightarrow 3$	DN	DN	0.8227	18	14
(**)	$1 \rightarrow 4$	$2 \rightarrow 3$	$2 \rightarrow 3$	DN	0.945	18.8	14.6

(*) SM strategy when there is no benefit of repairing multiple components
(**) SM strategy when considering the benefits of repairing multiple components
DN Do nothing

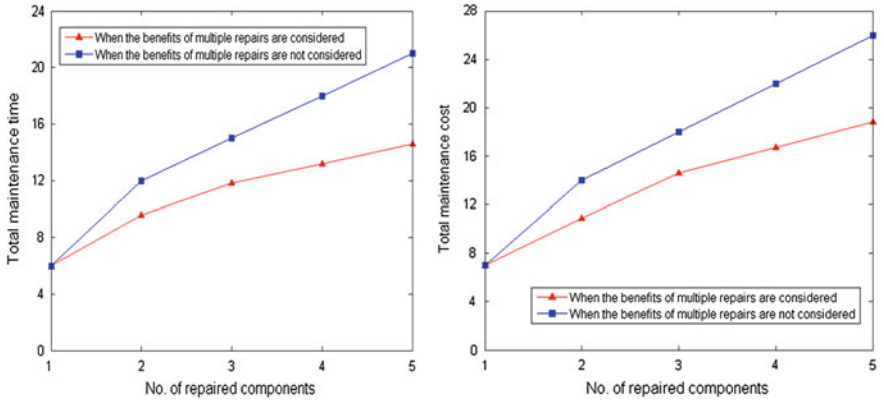


Fig. 5 Total maintenance time (**a**) and cost (**b**) of repairing multiple components

7 Conclusion

This paper studies the selective maintenance model for multi-state series-parallel system considering the benefits of repairing multiple components simultaneously. The model is closer to practical applications when the maintenance resources for repairing components are dependent. Both time and cost savings can be assured when several components are maintained in a selective maintenance strategy. The advantages of multiple repairs are analyzed with regards to two types of time and cost savings: (i) the share of setting up and (ii) the additional saving due to repairing multiple identical components in the multistate series-parallel systems. A non-linear integer optimization model is developed to maximize the system reliability at the next operating mission subjected to both time and cost constraints. Genetic Algorithms is used to solve the optimization models.

In general, the selective maintenance problem in this paper can help the maintenance manager determine the best maintenance strategy to get a reliable system and locate the maintenance resource effectively. The illustrative example shows that more maintenance actions can be performed and higher reliability of the system can be achieved if the benefits of repairing multiple components are addressed.

Acknowledgment This research is supported by Natural Sciences and Engineering Research Council of Canada (NSERC) and Vietnam International Education Development (VIED).

References

1. Rice WF, Cassady CR, Nachlas JA (1988) Optimal maintenance plans under limited maintenance time. In: Industrial engineering research conference
2. Cassady CR, Pohl EA, Murdock WP (2001) Selective maintenance modeling for industrial systems. J Qual Maint Eng 7(2):104–117
3. Schneider K, Cassady CR (2004) Fleet Performance under Selective Maintenance, RAMS
4. Maillart LM, Cassady CR, Rainwater C, Schneider K (2009) Selective maintenance decision-making over extended planning horizons. IEEE Trans Reliab 58(3):462–469
5. Pandey M, Zuo MJ, Moghaddass R (2012) Selective maintenance for binary systems using age-based imperfect repair model. In: International conference on quality, reliability, risk, maintenance, and safety (QR2MSE)
6. Chen C, Meng MQ-H, Zuo MJ (1999) Selective maintenance optimization for multi-state systems. In: Proceedings of the IEEE Canadian conference on electrical and computer engineering
7. Liu Y, Huang HZ (2010) Optimal Selective maintenance strategy for multi-state systems under imperfect maintenance. IEEE Trans Reliab 59(2):356–367
8. Pandey M, Zuo MJ, Moghaddass R (2013) Selective maintenance for multistate system with multistate components. IIE Trans 45(11):1221−1234
9. Maaroufi G, Chelbi A, Rezg N (2012) A selective maintenance policy for multi-component systems with stochastic and economic dependence. In: 9th international conference of modelling, optimization and simulation
10. Nourelfath M, Chatelet E (2012) Integrating production, inventory and maintenance planning for a parallel system with dependent components. Reliab Eng Syst Saf 101:59–66
11. Sivanandam SN, Deepa SN (2010) Introduction to genetic algorithms. Springer, Berlin

The Development of ISO 55000 Series Standards

M.R. Hodkiewicz

Abstract The launch of a set of three ISO Standards for Asset Management in 2014 represents a step change for the AM community. One of the Standards, ISO 55001 seeks to be applicable to all asset owners from Art Galleries to the Defence Force and will be used for certification. With over 25 participating countries there is expected to be a large number of organisations, regulators, and governments who look to these Standards for asset management and certification guidance. This paper provides an overview of the ISO 55001 standard, its relationship to the other standards in the set as well as identifying differences to the PAS55 Specification which preceded it. It describes the development process and identifies areas of major debate for the committee over the 3 years. These debates often relate to areas that lack theory and quantifiable evidence and hence represent opportunities for research, these are identified. The paper will be of value to those who are not familiar with the ISO process and wish to understand why the Standards have the content they do and how they will contribute to the future of Asset Management.

Keywords Standard · ISO55001 · Asset management · PAS55

1 Introduction

All organisations own assets, the physical assets are used to produce products and deliver services. For decades senior managers have asked if this is being done efficiently and effectively. Questions such as do we have the right assets? Are they delivering what we need now and will they do so in the future? What if they fail? What do they cost us to operate? How will they be impacted by new technology and changing external practices and events? The practice of asset management (AM) seeks to ensure that how physical assets are used and how decisions concerning

M.R. Hodkiewicz (✉)
University of Western Australia, Crawley, Australia
e-mail: melinda.hodkiewicz@uwa.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_37

assets are made are aligned to the needs of the organisation. Debate about AM practice has intensified in the last decade as organisations have realised both the risks and opportunities presented by how they manage their physical assets. This has led to the development of technical societies and sector bodies aimed at developing ideas around best practice in AM and from these Standards have emerged. This paper describes the development of such a set of Standards, the ISO 55000 series and provides insight into some of the issues that challenged the group responsible for its development as well as highlighting opportunities for the future.

2 Development Process

Organisations have been managing assets for decades and by the early 2000s there was a wealth of practice, process and ideas around the subject of asset management. However sharing this knowledge across sectors (power, water, roads, resource, local government etc.) was complicated by the use of different terms, definitions and process compounded by technical societies, research and sector groups promoting their view of best practice. Academics and consultants who moved across the sectors could see that in many cases the differences were quite artificial and there was much to learn from individual sectors. This thinking led to the development of the PAS 55 specification in 2004 under the leadership of the Institute of Asset Management (IAM). PAS 55 galvanised the asset management community especially the infrastructure sector in Europe and the increased interest and input resulted in a substantial revision in 2008 [1, 2]. However the re-issue had been done as a British Standards Institute (BSI) Specification and with global interest in the contents BSI initiated moves to translate PAS 55 into an ISO Standard.

In 2009 BSI issued a New Work Item Proposal (NWIP) for the development of 3 standards for Asset Management. Draft standards using the PAS 55 text with ISO cover sheets were distributed to Standards bodies around the world for comment and to assess interest in participation. The NWIP meeting was held in London in 2010. Two issues at this meeting fundamentally changed the pathway of development of the Standard. The first is that the ISO Technical Management Board had been working on a new template for their management system standards (MMS) and wanted the new AM Standard to use it. The template, called the JTCG text, is intended to assist organisations that use multiple ISO MMS to streamline their processes through use of a common structure in the document and, where possible, common text. This common text only applies to the 'Requirements' document in the MMS set. The need to use the JTCG standard text (known as blue text) as the base document provided an opportunity for a fresh look at an AM Standard, free of the requirement to adhere to the original PAS 55 text. The second issue, a consequence of the first, was that the non-IAM groups and sectors involved in the NWIP then had the opportunity to bring their ideas, terms and definitions to the table. Examples of relevant material include the International Infrastructure Management Manual [3], Road Sector asset management texts, and materials from other

professional bodies. In practice despite the use of the JTCG text and all the new input there much is common between PAS55 and ISO55001 but the journey has been useful in ensuring that all countries feel they have had a say in its development.

Development of the three standards was approved in 2010 and subsequently meetings were held at about 9 month intervals. Global interest has continued to grow and at the 4th meeting in Prague 2012 27 countries participated. The addition of new national groups and asset sectors has contributed to a dynamic development environment. The standards are due for issue in the 2nd quarter of 2014 and will be accompanied by a standard ISO 17021-5 [4] setting out the requirements for competence of personnel involved in the certification process.

It is important to remember that the main aim of Standards is to promote good practice, not to be there purely for compliance. As mentioned earlier infrastructure operators who were early adopters of PAS55 are likely to embrace ISO55001 particularly if it is a requirement or recommendation from their regulator. These include government and privatised groups in the gas, water, electricity and transport sectors and companies in these sectors are watching developments carefully. There is also possibility that central/federal governments will impose a requirement to comply with the Standards for other levels of government (e.g. state and local) under their control (insert ref PWs paper). The ISO 55000 series is intended to cover "all" assets so it will be interesting to see if additional sectors adopt it (for example, facilities management, real estate, manufacturing, defence).

3 Overview of the Documents

There are three documents as follows:

- ISO 55000 Asset Management – Overview, Principles and Terminology,
- ISO 55001 Asset Management—Management System—Requirements,
- ISO 55002 Asset Management—Management System—Guidelines on the application of ISO 55001.

ISO 55000 Standard describes the major dimensions assets, asset management (AM) and the AM Management System (AMMS) as well as presenting the principles of AM and terms and definitions. It includes a section describing the potential benefits of AM. The Standard defines an asset as "something that has potential or actual value to an organization" [5]. The value depends on the context of the organization and its stakeholders. Value can be tangible or intangible, financial or non-financial. Asset management is the "set of coordinated activities that an organization uses to realize value from assets" acknowledging that realization of value normally involves a balancing of costs, risks, opportunities and performance benefits" [5]. The AMMS is the "set of interrelated or interacting elements of an organization that establish AM policies and objectives, and the processes needed to

achieve those objectives". The AMMS essentially defines the business processes used by those in the organisation working on or with the assets.

The ISO 55001 Standard specifies the requirements for the establishment, implementation, maintenance and improvement of a management system for asset management. It does this through ~67 shall statements (many with multiple clauses). An example of a requirement is *The organization shall determine: the stakeholders that are relevant to the asset management system; the requirements of these stakeholders; the requirements for recording financial and non-financial information, and for reporting on it both internally and externally; the criteria for asset management decision making, and the requirements for alignment of financial and non-financial terminology throughout the organization.* In this example, the text of from the start of the requirement through to the second clause is "blue" text which is common to other ISO MSS, and the last three clauses (underlined) are specific to asset management and were added by the committee working party.

Contents of the standard cover core management system elements such as understanding the needs of the organisation and its stakeholders, leadership expectations, policy, roles and responsibilities, planning, setting objectives, and the requirements for supporting functions. These supporting functions include resourcing, competencies, awareness and communication, information and documentation requirements, operational planning and control, management of change, outsourcing, performance evaluation, management review and improvement processes.

ISO 55002 Standard follows the same layout of contents as in ISO 55001 but has no "requirements", instead focussing on providing further clarification of the intent of the requirements clauses in 55001.

4 Differences Between PAS 55-1 and ISO 55001

Both PAS 55-1 and ISO 55001 set out requirements and can be used for certification. Although the documents are similar in intent and much of the contents, they have different layouts and there are some differences. These are summarised below.

4.1 Scope

PAS 55 was primarily focussed on the management of physical assets and asset systems whereas ISO 55001 has a wider remit and the Standard says it can be applied to all types of assets; however it is does qualify this by saying that it is particularly intended to be used for managing physical assets.

4.2 AM Strategy and AM Strategic Plan

There is no mention of AM Strategy in ISO 55001 whereas there was an entire section in PAS 55-1. The ISO committee has decided to clearly differentiate what goes in an AM Plan and what is part of the Strategic Plan for AM. The AM Strategic Plan is intended to align with organisation's strategic planning process, be integrated with the development of the AM objectives, and focus on the role of the AMMS in delivering on these objectives. In contrast in PAS 55 the boundary of the strategy document between a strategic and operational focus was somewhat unclear as clauses required organisations to "identify the function(s), performance and condition of existing asset systems and critical assets; state the desired future function(s), performance and condition of existing and new asset systems and critical assets, on timescales aligned to those of the organizational strategic plan; clearly state the approach and principal methods by which assets and asset systems will be managed." These are details at the asset level and in the ISO document all of this information is contained in the AM Plans.

4.3 Financial Information and Requirements

There is only one reference to the word 'financial' in the main body of the PAS 55-1 text and that is with respect to asset management objectives needing to take account of financial requirements. ISO 55001 makes two important additions, one with respect to financial information and the other to ensure that the financial implications of AM planning are clearly articulated. With respect to information the requirements for recording financial information and for aligning financial terminology across the organisation need to be defined and traceability ensured. As users of computerised maintenance management systems will know, there can be challenges in obtaining cost data at the asset level particularly in organisations that don't use activity based costing. On reporting on performance there is also requirement to report financial information at the asset level, asset management performance and with respect to the effectiveness of the AMMS. The last requirement implies that the cost of managing the AMMS will need to be determined.

4.4 Contingency Planning

PAS55 has a separate section on contingency planning. There is no specific mention of contingency planning in ISO 55001. The assumption being that this will be covered in the risk controls put in place as part of an asset risk management program.

4.5 Risk Management

ISO 55001 assumes that the organisation will be using the ISO 31000 risk management framework (or equivalent) which specifies all the steps in the risk management process. It therefore does not spell these out in the 55001 text. The approach is to identify what needs risk management, for example, the risk of not meeting asset management objective and require the organisation to ensure that such risks are managed within its risk management processes using the organisation's risk management approach.

PAS 55 has a separate section on risk management process setting out requirements for the risk management methodology and risk identification and assessment. The risk evaluation and control steps covered in ISO 31000 are not however covered in detail in PAS 55. The PAS 55 document also specifies requirements for asset risk information.

4.6 Information Management

While the requirements around what information to collect are broadly similar across the two documents there is more detail required in the IS0 55001 on data quality attributes. The organisation will need to ensure the documented information has appropriate identification, description, format, and has been reviewed for suitability and adequacy. Data quality attributes for availability and protection are also addressed. There is also a requirement for information on the asset performance monitoring, measurement, analysis and evaluation, nonconformities or incident and any subsequent actions taken, and the results of any corrective action. This will likely create a wealth of asset data, potentially of higher quality than is currently available.

4.7 Competence

There are two major differences around competence. PAS 55 focussed on the identification of competency requirements, the development of plans, and provision of training. ISO 55001 extends this by adding requirements to ensure that people are competent, and where education/training/experience are provided to evaluate the effectiveness of this with respect to the competence required. This implies there will need to be programs that assess competence. The other difference is the onus on ensuring competence of outsourced personnel is moved from the needing to ensure that outsource providers have arrangements in place to demonstrate their staff are competent (the PAS 55 requirement) to having to ensure that outsourced resources meet the same competence requirements as for internal resources (in ISO 55001).

5 Areas of Debate During ISO 55000 Series Development

This ISO process which brought together people from different asset sectors, nationalities, asset management background, was one involving of lively debate. This primarily occurred in intensive 1 week sessions involving 50–70 people, all experts in their fields, in a hot-house environment with a deadline each time to produce a new document having addressed all the comments made by national and liaison bodies on the previous document. ISO 55001 was the easiest document to work on because so much of the text is proscribed by the JTCG format. Decisions had to be made on what to add. This was done by sub-groups coordinated by the ISO 55001/55002 working group chair. ISO 55000 on the other hand had no proscribed format and many ideas about what should be in and how it should be represented. It was developed by a separate working group under a different chair. Although there were many areas of debate, some of the highlights and the compromises reached are discussed below.

5.1 Scope

As mentioned earlier the ISO 55001 reports in the scope section that it "can be applied to all types of assets, by all types and sizes of organisations". The compromise for committee members who are deeply uneasy about the potential breadth of "all" which could encompass intangible assets such as brands and reputation, financial and human assets has been to add a note to say that it is "intended to be used for managing assets but this does not limit its application to other asset types." One of the concerns raised is that most committee members come from a background of managing physical assets and have little experience in non-physical asset management from which to judge if the standard is indeed applicable to these.

5.2 Applicability

There is concern about the ability of smaller organisations for example local councils and small infrastructure organisations that they will have the resources to implement the requirements, especially since the ISO 55001 standard stipulates that the "requirements of the Standard should be applied in its entirety". To manage these concerns the clause "to the assets determined by the organisation" has been added to the sentence. In practice this means that a local council could choose specific assets such as parks to apply the standard to but not others such as parks and public artwork.

5.3 Asset Life and Life Cycle Definitions

The concept of asset life has been hotly debated. In ISO 5500 "Asset life" is defined as the period from conception to end-of-life for an asset. An "asset life cycle" includes all the stages that an asset experiences over its "asset life". In some organisations assets are purchased, used and disposed of while others such as road have no realistic end of life stage. It has been left up to the organisation to determine how to define and name each stage of the asset life cycle. Another issue is that asset's life cycle does not necessarily coincide with the period over which any one organization holds responsibility for the asset. For example with privatisation of rail lines, the responsibility for the asset changed but it didn't impact where the rail was in its life cycle. The Standard recognises that an asset can hold value to one or more organizations over its life. A new term, the AM life cycle was introduced. This describes the period of time that the assets are under the control of the organization. The asset management life cycle begins with the commencement of responsibility period for the organization. Economic life and technical life are mentioned in ISO 55002 with respect to AM Plan development.

5.4 Figures

There is widespread agreement that figures will help particularly those from non-English speaking backgrounds to read the standard. However there has been no agreement on what the figures should be. One of the main points of concern is to clarify the difference between asset management (what is done to/with the assets) and the asset management system (the processes that determine how it is done, when and by whom). Many of the proposed figures are essentially a pictogram of the table of contents in blocks with various connecting arrows and there are many ways this can be done. There is little published work based on evidence about what the key input-output relationships are and how AM is actually done in organisations.

5.5 Financial Management

A finance sub-committee was established to coordinate the integration of financial management and asset management through the standards. While this has resulted in many welcome developments such as the requirement to determine the financial implications of the AM plan and traceability between financial and technical asset data there are those who feel that it does not go far enough. There is an argument to say that doing all the plans and then not getting the funds to action them is a waste of resources, but there are others concerned that the maturity of many organisations with respect to AM plans and the lack of integration between their accounting and asset systems means that the Standard cannot require too much too soon.

5.6 Asset Management Principles

These have been through much iteration as there has been no research such as that done by the quality management community [6, 7] on what different organisations identify as being core to successful asset management. The challenge facing the asset management community, similar to those in the risk, quality, safety, and energy management communities is to decide whether to focus on principles which are core to all management systems such as leadership or to focus on principles which are specific to their sector. An example would be a principle associated with an issue like the need to balance cost, risk and performance over the asset management life cycle, which is specific to asset management.

6 Opportunities for Research

There are three sets of research questions identified here, one set are at a macro level and look at asset management as a discipline, the next is about modelling and prediction and the final one looks at the opportunities provided by more and better data on assets, how we manage them and how they perform.

6.1 How to Measure AM and Its Impact on Organisational Performance

Each sector has historically developed its own definitions, frameworks, models, and body of knowledge for AM and vigorously promoted these to their own communities. Although there is an emerging consensus through ISO 55001 development process about which factors are important to AM and organisational success, these have not been tested using validated, statistical methods. Research questions include:

1. What are the factors that characterise AM practice?
2. Does AM practice delivers superior AM and/or organisational outcomes?
3. Will ISO 55001 Certification deliver improved organisational performance?

To answer question 1 there will need to be surveys across organisations in the same, and in some cases, different sectors to identify common factors as well as factors that are sector specific. To answer questions 2 there is a work to do to determine how to link practice and activities to AM and organisational outcomes. This is about having demonstrable evidence that “AM performance is positively associated with committed leadership” and “AM organisations demonstrate better safety performance than non-AM organisations”. Question 3 can only be addressed once ISO 55001 is issued but a look back at the history of the quality management

movement shows that there are reputable studies both for and against the hypothesis that "Organisations that certify outperform organisations that do not" and "The benefits of certification outweigh the costs" [8, 9].

This research work involves support from a number of organisations (typically more than 20 would need to participate) and willingness to participate in surveys and provide technical and financial information. The risk of not doing this is that the benefits of AM are unquantified and largely linked to the perceived benefits of a few companies and the marketing material of consultants. One of the challenges, not insurmountable, with this area of research is it is in the domain of the management scientists and uses statistics and processes that are not familiar to the technical and engineering community from which many asset management practitioners and researchers are drawn.

There is also a requirement in ISO 55001 to determine the effectiveness of the AM system, which leads to questions about how we would measure effectiveness as well as establish costs.

6.2 How to Predict Future Asset Management Requirements and Impacts

The second area of research centres on the emphasis in ISO 55001 on planning. Good planning is conditional on having understanding a range of realistically possible views of the future. This is enabled by modelling work that allows the organisation to predict potential outcomes and plan accordingly. There is currently a number of research and industry groups that look at infrastructure policy in the various sectors e.g. energy and transport who seek to include more representative asset performance and behaviour assumptions in their models. Within organisations there are also modelling groups that seek to understand the range of impacts of various external policy decisions and physical/environmental events (e.g. climate change) on their assets. Typically these models use socio-technical modelling and/or economic modelling approaches such as agent based modelling and serious gaming for the former and real options is an example of the latter approach.

6.3 Will More and Better Data Lead to New Insights in Asset Management

The final area of research is focussed on the expectation that more and better data will be collected on assets. This will allow the traditional engineering researchers interesting in remaining useful life and similar maintenance and reliability data models a much larger pool of data both on failures but also potentially on relevant influence variables and confounding factors. More data also opens the door to

algorithmic models that don't presuppose a relationship between inputs and outputs but use machine learning tools to develop an understanding of patterns of behaviour [10]. Work in this area, topically called 'Big Data' requires a high level of competence in statistical methods.

One of the challenges inherent in the three questions posed here is that they require tools and techniques that are not core to the civil, mechanical, electrical, and chemical engineering disciplines from which the majority of academics working in AM are drawn from. Does the future of AM as a discipline depend on the research community drawing in experts in these fields on an ad hoc basis or are these tools we should be adding to our toolbox? What does the AM research community look like in 2025 is an interesting question to consider.

7 Conclusion

This is essentially an optimistic paper. It recognises that AM as a discipline and professional practice has come a long way since 2004 and that the launch of the ISO 55000 series Standards will be a major milestone on that journey. In describing elements of the journey and particularly some of the challenges, the paper highlight the compromises necessary as part of what is a complicated process. It also identifies a number of areas where theory and evidence are necessary to improve the quality and content of future debates and new opportunities for research.

Acknowledgments The author would like to thank Professor Paulien Herder for the opportunity to visit TU Delft, during which time this paper was written. She would also like to acknowledge and thank members of the Standards Australia MB19 group involved in developing the ISO 55000 standards who have helped to shape and inform her views over the last 3 years.

References

1. BSI (2008) Asset management part 1: specification for the optimised management of physical infrastructure assets. British Standards Institute, England
2. BSI (2008) Asset management part 2: guidelines for the application of PAS 55-1. British Standards Institute, England
3. NAMS (2006) International infrastructure management manual. Thames, New Zealand: NAMS
4. ISO (2013 (expected)). ISO/IEC 17021 Conformity assessment—Part 5: Competence requirements for the certification of asset management systems. Switzerland: ISO/IEC
5. ISO (2012) ISO/PC 251 ISO 55000 DIS Asset management—Overview, principles and terminology International Standards Organization
6. Flynn BB, Schroeder RG, Sakakibara S (1994) A framework for quality management research and an associated measurement instrument. J Oper Manag 11(4):339–366
7. Garvin DA (1988) Managing quality. the strategic and competitive edge. The Free Press, New York

8. Chow-Chua C, Goh M, Tan BW (2003) Does ISO certification improve business performance? Int J Qual Reliab Manag 20(8/9):936–953
9. Terziovski M, Power D, Sohal AS (2003) The longitudinal effects of the ISO 9000 certification process on business performance. Eur J Oper Res 146(3):580–595
10. Breiman L (2001) Statistical modeling: the two cultures. Stat Sci 16(3):199–231

A Preventive Maintenance Model for Linear Consecutive k-Out-of-n: F Systems with Dependent Components

W. Wang, F. Zhao, R. Peng and L. Guo

Abstract Preventive maintenance (PM) is an important maintenance action to prevent from the occurrence of failures. This paper presents a PM policy for a linear consecutive k-out-of-n: F system that fails if and only if at least k consecutive components fail. The failure rate of a component depends on the state of the adjacent components since the load of a component may increase when its adjacent component fails. The failed components in the system are replaced with new ones either at the time of system failure or at the time of PM whichever comes first. A failure sequence diagram is presented to establish the reliability model of the system. Based on the reliability model, the optimal preventive maintenance strategy which minimizes the long-run expected system cost per unit time is studied with an illustrative example.

1 Introduction

A consecutive k-out-of-n: F (G) system, C(k, n; F/G) system, consists of a sequence of n ordered components along a line or a circle such that the system fails (works) if and only if at least k consecutive components in the system fail (work). Such a system was first introduced by Kontoleon in 1980 [1] and has been used to model telecommunications, oil pipelines, vacuum systems in accelerators,computer ring networks and space relay stations [2]. Many research works have been reported to

W. Wang (✉) · F. Zhao · R. Peng · L. Guo
University of Science and Technology Beijing, 100083 Beijing, China
e-mail: wangwb@ustb.edu.cn

F. Zhao
e-mail: zhaofei.19841027@163.com

R. Peng
e-mail: pengrui1988@gmail.com

L. Guo
e-mail: guoli@manage.ustb.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_38

deal with this type of system to derive the algorithms and bounds for reliability characteristics [3–7]. These papers mostly assume that the components in the system are identical and s-independent [3–5]. In practice, components may have different characteristics and depend on each other due to load-sharing or other reasons. Consecutive k-out-of-n systems with heterogeneous components were studied in [6]. However, the component failures are assumed to be independent.

Aki and Hirano investigated the lifetime distributions of C(k,n;F) systems with two types of dependence among system components, one was based on the notion of sequential order statistics and the other considered the dependence of the component failure rate on the state of the left adjacent component [8]. Some other works also studied the reliability of consecutive k-out-of-n systems with dependent components [9–14]. Lam and Zhang studied a repairable C(2,n;F) system with Markov dependence where the failure rate of a component depends upon, and only upon, the state of the preceding component and the lifetime and repair time of components are exponential distributed with various parameters [10]. Xiao et al. presented a model of C(k,n;F) repairable systems with non-exponential repair time distribution and (k-1)-step Markov dependence [11]. Villén-Altamirano analyzed non-Markov C(k,n;F) repairable systems by revising the model in [11] with the assumption that the component lifetimes follow a general distribution [12]. Eryılmaz studied reliability properties of C(k,n;F/G) systems with arbitrarily dependent components [13]. Yun et al. reported the optimal assignment for a circular C(k,n;F) system with (k-1)-step Markov dependence [14]. The focus of all these literatures is on reliability modeling and optimal system design, and none of them have considered the optimal preventive maintenance strategy for consecutive k-out-of-n systems with dependent components.

This paper considers a C(k,n;F) system with dependent components, which is motivated by the real practice in the steelmaking industry. Billets are transferred from the withdrawal and straightening system "A" to the tilting manipulator "B" through a roller system, which consists of n chains equally spaced between A and B, as shown in Fig. 1. Whenever the length of the consecutive failed chains is larger than the length of the billet "L", the roller system is not able to transfer the billet successfully and is regarded as failed. Therefore the system fails if and only if at least consecutive k chains fail, where $k = \arg\min\{kl/L\}$ and l is the length of each

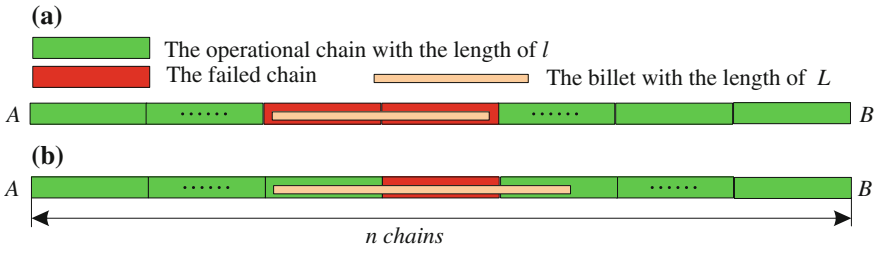


Fig. 1 The roller system in the steelmaking industry

chain. Further, the failure of a chain results in more load shared by its adjacent chains and increases their failure rates.

The outline of this paper is organized as follows. Model assumptions and the notations are given in Sect. 2. Section 3 proposes a failure sequence diagram for reliability modeling of the C(k, n; F) system. Based on the reliability model, the optimal preventive maintenance strategy is studied for a case with $k = 2$ and $n = 4$ in Sect. 4. Section 5 concludes this paper and points out some future researches.

2 Modeling Assumptions and Notation

We first specify the assumptions and the notation that will be used hereafter.

(1) The system consists of n components arranged linearly and the system fails whenever at least k consecutive components fail.
(2) The components and system are either operational or failed.
(3) If one component in the system fails, the failure rates of its adjacent components increase as more loads are shared by them.
(4) The probability that two or more than components in the system fail simultaneously is negligible.
(5) The components in the system have an exponential distribution with failure rate$\lambda_0$if the adjacent components are in the operational state. However, the failure rate of a component changes toλ_1 if one of its adjacent components fails, andλ_2 if both adjacent components fail ($\lambda_0 < \lambda_1 < \lambda_2$).
(6) Once the system fails, the failed components in the system will be replaced with new ones immediately.
(7) The system is repaired preventively every planned interval with the failed components replaced by new ones.
(8) It is considered that there are always enough spare parts available on hand.

Notation

k	Minimum number of consecutive failed components which lead to a failure of system
n	Number of components in the system
λ_j	Possible failure rate of working components where j = 0,1,2
f_{i,λ_j}	Probability density function *(pdf)* of component i with failure rate λ_j, where i = 1,2,… n
R_{i,λ_j}	Reliability function of component i with failure rate λ_j
T	Preventive maintenance interval
t_i	Time to failure of component i
c_r	Replacement cost of a failed component
c_p	Downtime cost due to PM
c_f	Downtime cost due to failure
$EC(T)$	Expected renewal cycle cost
$EL(T)$	Expected renewal cycle length
$C(T)$	Long-run expected cost per unit time

3 The Failure Sequence Diagram

In this section, a failure sequence diagram (FSD) is proposed for reliability modeling of the $C(k, n; F)$ system with dependent components. The procedures are given as follows:

(1) Start the FSD construction with the representation of the entire system "Ⓢ", , see Fig. 2. The node *0* in Fig. 2 shows that there are no failed components in the entire system. The component node "ⓘ" represents that the component i is the first failed component in the system, where i = 1, 2, . . . ,n.

(2) According to the characteristics of $C(k, n; F)$ systems that the system fails if and only if k consecutive components in the system fail, we need to judge if the number of the consecutive failed components in any path of Fig. 2 is greater than or equal to k. If the condition is satisfied, then it indicates that the system fails and there is no need to add any more nodes to the path. Otherwise, for the component node "ⓘ" (i = 1,2, . . . ,n) in Fig. 2, construct the FSD of "ⓘ" by adding node *0* and the nodes for un-failed components as the child nodes of node "ⓘ". For example, the child nodes of node "①" are node *0* and component nodes *2, ... ,n*; and the child nodes of node "ⓝ" are node 0 and component nodes 1, ... ,$n - 1$. Figure 3 gives only the FSD of "ⓘ" ($1 < i < n$) for simplicity. A path from S to i to j means that the first two failed components in the system are component i and j.

(3) Once the ith level for the FSD has been constructed, the $(i + 1)$th level for the FSD can be constructed similarly to step 2. The construction of the FSD stops until all the paths end either at node 0 or due to the failure of the system.

The path ending at node *0* implies that the system is operational as the number of consecutive failed components is less than k. Other paths show that the system fails as there are at least k consecutive failed components in the system.

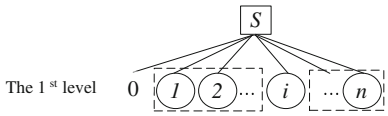


Fig. 2 The FSD representation of the entire system "Ⓢ"

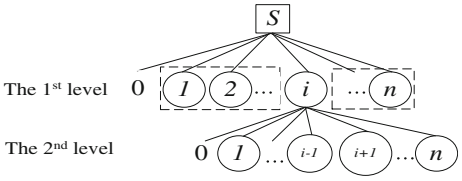


Fig. 3 Illustration of the FSD representation for component node "ⓘ"

4 Numerical Example

4.1 The PM Model of Case k = 2 and n = 4

We take the long-run expected cost per unit time as an objective function to find the optimal PM interval. From assumptions (6) and (7), it can be concluded that the system is renewed either at time of PM or at the failure of the system. In order to derive the long-run expected cost per unit time, all possible renewal possibilities need to be formulated. The FSD of $C(2,4;F)$ systems is given in Fig. 4.

(1) Probability of a PM renewal

Since "0" representing the terminal events that the system is operation at any time, so from Fig. 4 we can see that all possible renewal events due to PM are E_{S0}, E_{S10}, E_{S130}, E_{S140}, E_{S20}, E_{S240}, E_{S30}, E_{S310}, E_{S40}, E_{S410}, and E_{S420}. Due to the symmetry of components in the structure of system, we just need to count the possibilities for the event set $\{E_{S0}, E_{S10}, E_{S130}, E_{S140}, E_{S20}, E_{S240}\}$. Figure 5 represents the change of failure rate for the components in the system when the event E_{S130} happens.

The system can still operate after components 1 and 3 fail from assumption (1) so that a PM renewal is done at the time of PM T to replace the failed components. Since component 1 fails firstly and then component 3 fails, in terms of assumption (5), the failure rate of the component 2 is λ_1 after t_1 but λ_2 after t_3 and the failure rate of the component 4 is λ_1 after t_3. The probability of event E_{S13} is given by

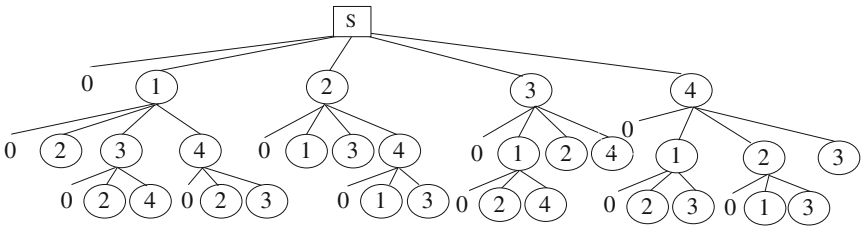


Fig. 4 The FSD representation for consecutive 2-out-of-4: F systems

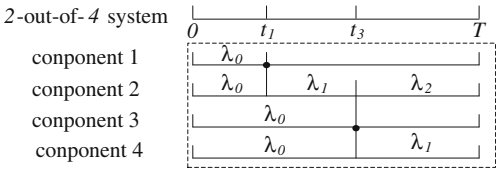


Fig. 5 Illustration of event E_{S130}, where ● denotes the component fails and $0 < t_1 < t_3 < T$

$$P(E_{S130}) = \int_0^T \int_{t_1}^T f_{1,\lambda_0}(t_1) R_{2,\lambda_0}(t_1) R_{2,\lambda_1}(t_3 - t_1) R_{2,\lambda_2}(T - t_3) f_{3,\lambda_0}(t_3) R_{4,\lambda_0}(t_3) R_{4,\lambda_1} (T - t_3) dt_3 dt_1. \tag{1}$$

Similarly to Eq (1), the possibilities of other PM renewal events are given as

$$P(E_{S0}) = \prod_{i=1}^{4} R_{i,\lambda_0}(T), \tag{2}$$

$$P(E_{S10}) = \int_0^T f_{1,\lambda_0}(t_1) R_{2,\lambda_0}(t_1) R_{2,\lambda_1}(T - t_1) R_{3,\lambda_0}(T) R_{4,\lambda_0}(T) dt_1, \tag{3}$$

$$P(E_{S140}) = \int_0^T \int_{t_1}^T f_{1,\lambda_0}(t_1) R_{2,\lambda_0}(t_1) R_{2,\lambda_1}(T - t_1) R_{3,\lambda_0}(t_4) R_{3,\lambda_1}(T - t_4) f_{4,\lambda_0}(t_4) dt_4 dt_1, \tag{4}$$

$$P(E_{S20}) = \int_0^T R_{1,\lambda_0}(t_2) R_{1,\lambda_1}(T - t_2) f_{2,\lambda_0}(t_2) R_{3,\lambda_0}(t_2) R_{3,\lambda_1}(T - t_2) R_{4,\lambda_0}(T) dt_2, \tag{5}$$

$$P(E_{S240}) = \int_0^T \int_{t_2}^T R_{1,\lambda_0}(t_2) R_{1,\lambda_1}(T - t_2) f_{2,\lambda_0}(t_2) R_{3,\lambda_0}(t_2) R_{3,\lambda_1}(t_4 - t_2) R_{3,\lambda_2} (T - t_4) f_{4,\lambda_0}(t_4) dt_4 dt_2. \tag{6}$$

(2) Probability of a failure renewal

It can be seen from Fig. 4 that all renewal events due to a functional failure are E_{S12}, E_{S132}, E_{S134}, E_{S142}, E_{S143}, E_{S21}, E_{S23}, E_{S241}, E_{S243}, E_{S312}, E_{S314}, E_{S32}, E_{S34}, E_{S412}, E_{S413}, E_{S421}, E_{S423}, and E_{S43}. Due to the symmetry of components in the structure of system, we only list the following renewal probabilities.

$$P(E_{S12}) = \int_0^T \int_{t_1}^T f_{1,\lambda_0}(t_1) R_{2,\lambda_0}(t_1) f_{2,\lambda_1}(t_2) R_{3,\lambda_0}(t_2) R_{4,\lambda_0}(t_2) dt_2 dt_1. \tag{7}$$

$$P(E_{S132}) = \int_0^T \int_{t_1}^T \int_{t_3}^T f_{1,\lambda_0}(t_1) R_{2,\lambda_0}(t_1) R_{2,\lambda_1}(t_3 - t_1) f_{2,\lambda_2}(t_2) f_{3,\lambda_0}(t_3) R_{4,\lambda_0}(t_3) R_{4,\lambda_1} \times (t_2 - t_3) dt_2 dt_3 dt_1. \tag{8}$$

$$P(E_{S134}) = \int_0^T \int_{t_1}^T \int_{t_3}^T f_{1,\lambda_0}(t_1) R_{2,\lambda_0}(t_1) R_{2,\lambda_1}(t_3 - t_1) R_{2,\lambda_2} \times (t_4 - t_3) f_{3,\lambda_0}(t_3) R_{4,\lambda_0}(t_3) f_{4,\lambda_1}(t_4) dt_4 dt_3 dt_1. \tag{9}$$

$$P(E_{S142}) = \int_0^T \int_{t_1}^T \int_{t_4}^T f_{1,\lambda_0}(t_1) R_{2,\lambda_0}(t_1) f_{2,\lambda_1}(t_2) R_{3,\lambda_0}(t_4) R_{3,\lambda_1}(t_2 - t_4) f_{4,\lambda_0}(t_4) dt_2 dt_4 dt_1. \tag{10}$$

$$P(E_{S143}) = \int_0^T \int_{t_1}^T \int_{t_4}^T f_{1,\lambda_0}(t_1) R_{2,\lambda_0}(t_1) R_{2,\lambda_1}(t_3 - t_1) R_{3,\lambda_0}(t_4) f_{3,\lambda_1}(t_3) f_{4,\lambda_0}(t_4) dt_3 dt_4 dt_1. \tag{11}$$

$$P(E_{S21}) = \int_0^T \int_{t_2}^T R_{1,\lambda_0}(t_2) f_{1,\lambda_1}(t_1) f_{2,\lambda_0}(t_2) R_{3,\lambda_0}(t_2) R_{3,\lambda_1}(t_1 - t_2) R_{4,\lambda_0}(t_1) dt_1 dt_2. \tag{12}$$

$$P(E_{S23}) = \int_0^T \int_{t_2}^T R_{1,\lambda_0}(t_2) R_{1,\lambda_1}(t_3 - t_2) f_{2,\lambda_0}(t_2) R_{3,\lambda_0}(t_2) f_{3,\lambda_1}(t_3) R_{4,\lambda_0}(t_3) dt_3 dt_2. \tag{13}$$

$$P(E_{S241}) = \int_0^T \int_{t_2}^T \int_{t_4}^T R_{1,\lambda_0}(t_2) f_{1,\lambda_1}(t_1) f_{2,\lambda_0}(t_2) R_{3,\lambda_0}(t_2) R_{3,\lambda_1}(t_4 - t_2) R_{3,\lambda_2} \times (t_1 - t_4) f_{4,\lambda_0}(t_4) dt_1 dt_4 dt_2. \tag{14}$$

$$P(E_{S243}) = \int_0^T \int_{t_2}^T \int_{t_4}^T R_{1,\lambda_0}(t_2) R_{1,\lambda_1}(t_3 - t_2) f_{2,\lambda_0}(t_2) R_{3,\lambda_0}(t_2) R_{3,\lambda_1} \times (t_4 - t_2) f_{3,\lambda_2}(t_3) f_{4,\lambda_0}(t_4) dt_3 dt_4 dt_2. \tag{15}$$

(3) The expected renewal cycle cost and length

Using the above renewal possibilities, the expected costs for PM renewal and failure renewal are obtained as follows.

$$E(C_{pm}) = c_p P(E_S) + 2[(c_p + c_r)(P(E_{S10}) + P(E_{S20})) + (c_p + 2c_r)(P(E_{S130}) + P(E_{S140}) + P(E_{S240}))]. \tag{16}$$

$$E(C_f) = \begin{array}{l} 2[(c_f + 2c_r)(P(E_{S12}) + P(E_{S21}) + P(E_{S23})) + (c_f + 3c_r)(P(E_{S132}) \\ + P(E_{S134}) + P(E_{S142}) + P(E_{S143}) + P(E_{S241}) + P(E_{S243}))]. \end{array} \tag{17}$$

The expected renewal cycle length caused by a PM renewal $E(L_{pm})$is given by

$$E(L_{pm}) = T[P(E_{S0}) + 2[P(E_{S10}) + P(E_{S20}) + P(E_{S130}) + P(E_{S140}) + P(E_{S240})]]. \tag{18}$$

For the expected failure renewal cycle length, we need the pdfs of the failure events. For example for event E_{S12}, the probability that the system fails in *(0,z),* $z \in (0, T)$, is

$$P(t_f) = \int_0^z \int_{t_1}^z f_{1,\lambda_0}(t_1) R_{2,\lambda_0}(t_1) f_{2,\lambda_1}(t_2) R_{3,\lambda_0}(t_2) R_{4,\lambda_0}(t_2) dt_2 dt_1. \tag{19}$$

Then by differentiating Eq (19) with respect to z, we have the pdf of a failure at $t_f, t_f \in (z, z + dz)$for event E_{S12}, is given as

$$P(z < t_f < z + dz)/dz = -\lambda_1 e^{-(2\lambda_0 + \lambda_1)z}(e^{-2\lambda_0 z} - 1)/2. \tag{20}$$

Using Eq (20), we obtain the expected renewal cycle length of event E_{S12}

$$\int_0^T zP(z < t_f < z + dz) = \int_0^T z[-\lambda_1 e^{-(2\lambda_0 + \lambda_1)z}(e^{-2\lambda_0 z} - 1)/2]dz. \tag{21}$$

The similar derivation for the contribution to the expected renewal cycle length due to different failure event is omitted for simplicity. The expected renewal cycle length of a failure renewal$E(L_f)$ is obtained by summing up all contributions.

Based on the expected renewal cost and length, the renewal award theorem is used to model the objective function that determines the optimal PM interval by minimizing the long-run expected cost per unit time [15].

$$C(T) = \frac{EC(T)}{EL(T)} = \frac{E(C_{pm}) + E(C_f)}{E(L_{pm}) + E(L_f)}. \tag{22}$$

Table 1 The failure rate parameters and cost parameters

λ_0	λ_1	λ_2	c_r	c_p	c_f
0.030	0.125	0.205	10	100	1,000

4.2 Results for Optimal PM Interval

For the proposed model in Eq (22), we set firstly the failure rates for components and the cost parameters shown in Table 1.

Based on the parameters in Table 1, we obtain the optimal PM interval as 15 where the long-run expected cost per unit time is *5.404*0, see Fig. 6. The more frequent PM (smaller T) has a higher expected cost per unit time as the shorter PM interval increases the total cost of PM. However, the system fails easily if the PM interval is large such that the long-run expected cost per unit time increases because of the higher failure cost. We then further analyze the impact of the change of the failure cost c_f on the optimal PM interval while other parameters are fixed. Figure 7 shows that the optimal PM interval moves along the opposite direction of the change of the failure cost as we expect. When c_f = 700, the optimal PM interval is 16. When c_f = 1,500, the optimal PM interval moves to 14. Actually, when the failure causes a greater loss, it is advisable to shorten the PM interval in order to check the system more often to avoid unexpected system failure.

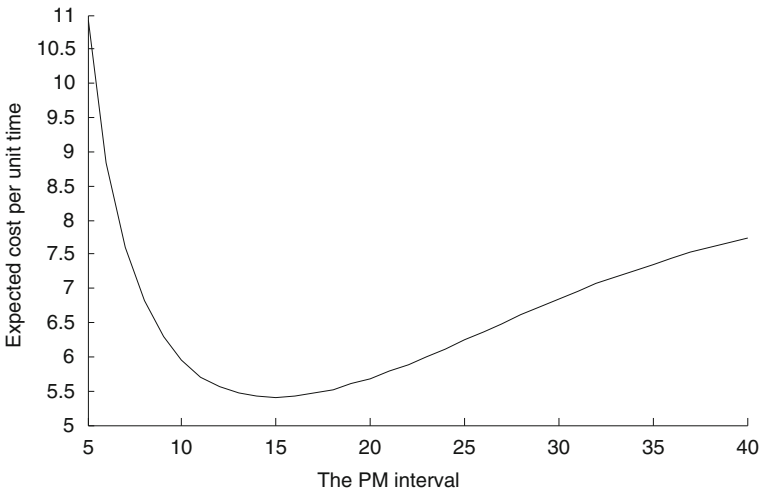


Fig. 6 Expected cost per unit time in terms of T

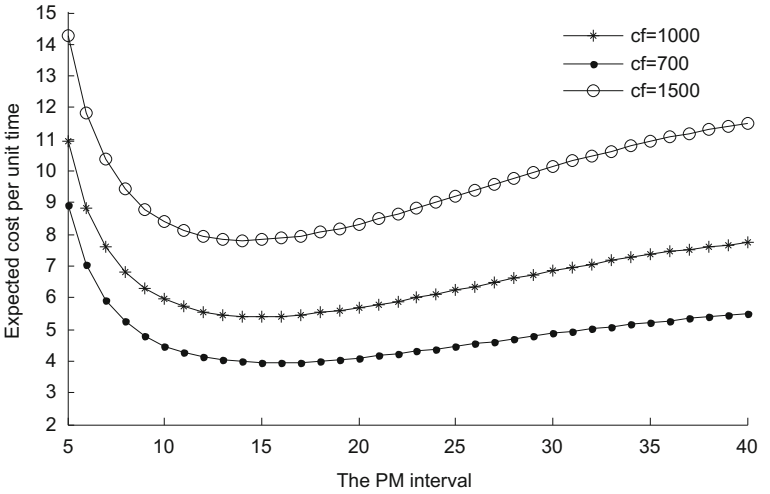


Fig. 7 Illustration of expected cost per unit time in terms of T when various failure costs

5 Conclusion

In this paper, we developed a failure sequence diagram for linear consecutive *k*-out-of-*n: F* systems with dependent components, which is inspired by the roller system in the steelmaking industry. Based on the failure sequence diagram, a PM model is established for a case with $k = 2$ and $n = 4$. Since PM is undertaken with a fixed interval and the system may fail before a scheduled PM takes place, there are two renewal scenarios, i.e., a PM renewal and a failure renewal. The model proposed aims to find the optimal PM interval which minimizes the long-run expected cost per unit time. The framework is shown by an illustrative example. Future studies will be developed towards proposing an algorithm for calculating the long-run expected cost per unit time of general linear consecutive k-out-of-n: F systems consisting of components with arbitrary lifetime distributions, and investigating the combined maintenance and spare parts optimization problem.

Acknowledgment The research report here was partially supported by the NSFC under grant number 71231001and 71301009, by the Fundamental Research Funds for the Central Universities of China under grant numbers FRF-SD-12-020A, FRF-SD-13-004B, FRF-MP-13-009A and FRF-TP-13-026A, and by the MOE PhD supervisor fund, 20120006110025.

References

1. Kontoleon JM (1980) Reliability determination of r-successive-out-of-n: F system. IEEE Trans Reliab R-29(5):437
2. Papastavridis SG, Koutras MV (1992) Consecutive k-out-of-n systems with maintenance. Ann Inst Statist Math 44(4):605–612

3. Chiang DT, Niu SC (1981) Reliability of Consecutive-k-out-of-n: F System. IEEE Trans Reliab R-30(1):87–89
4. Yam RCM, Zuo MJ, Zhang YL (2003) A method for evaluation of reliability indices for repairable circular consecutive-k-out-of-n: F systems. Reliab Eng Syst Safety 79(1):1–9
5. Yuan L, Cui ZD (2013) Reliability analysis for the consecutive-k-out-of-n: F system with repairmen taking multiple vacations. Appl Math Model 37(7):4685–4697
6. Salehi ET, Asadi M, Eryılmaz S (2011) Reliability analysis of consecutive k-out-of-n systems with non-identical components lifetimes. J Stat Plan Inf 141(8):2920–2932
7. Huang J, Zuo MJ, Fang Z (2003) Multi-State Consecutive-k-out-of-n Systems. IIE Trans 35 (6):527–534
8. Aki S, Hirano K (1997) Lifetime distributions of consecutive-k-out-of-n F systems. Proc 2nd World Cong Nonl Anal 30(1):555–562
9. Fu JC (1987) On Reliability of a large consecutive-k-out-of-n: F systems with k-1-step Markov dependence. IEEE Trans Reliab 36(1):75–77
10. Lam Y, Zhang YL (1999) Analysis of repairable consecutive-2-out-of-n: F systems with Markov dependence. Int J Syst Sci 30(12):1285–1295
11. Xiao G, Li Z, Li T (2007) Dependability estimation for non-Markov consecutive-k-out-of-n: F repairable systems by fast simulation. Reliab Eng Syst Safety 92(3):293–299
12. Villén-Altamirano J (2010) RESTART simulation of non-Markov consecutive-k-out-of-n: F repairable systems. Reliab Eng Syst Safety 95(3):247–254
13. Eryılmaz S (2009) Reliability properties of consecutive k-out-of-n systems of arbitrarily dependent components. Reliab Eng Syst Safety 94(2):350–356
14. Yun WY, Kim GR, Yamamoto H (2007) Economic design of a circular consecutive-k-out-of-n: F system with k-1-step Markov dependence. Reliab Eng Syst Safety 92(4):464–478
15. Ross SM (1983) Stochastic processes. Wiley, NewYork

Implementing Engineering Asset Management Standards (PAS-55) in Information Management Evaluation: Case Study in Hong Kong

Peter W. Tse, Jingjing Zhong and Samuel Fung

Abstract A number of facility management companies manage commercial buildings that are owned by the Government and private companies in Hong Kong. To ensure the performance in better quality, most of them have obtained different quality recognitions such as PAS 55 standards. However, in many companies, the organizations or building departments do not pay much attention on performing the criteria stated in the standards. They just follow the rules but there are no specific rules or procedures for performance evaluation. In this research, we investigated the implementation of the standards in asset information management that was performing in Hong Kong and conducted comprehensive analysis for current operating situation in asset information management. A structured questionnaire was designed and data was collected from 30 Operation and Maintenance (OM) departments for commercial buildings. These buildings are owned by the Government and private companies. The users of buildings include public services, commercial and banking business, residential tenants, industrial sectors and composite services in Hong Kong. The answers generated from the questionnaire reveal the real situation in implementing the standards of PAS-55, especially in asset information management. The answers show different performance levels for different kinds of buildings. The evaluation results point out that different types of buildings have different strategies in adopting the standards. From the data analysis, it reveals that there is a substantial gap exists between the adoption of standard implementing and the significance level claimed by the users in some cases. Hence, a gap analysis was conducted for these special cases. Furthermore, the relationship

P.W. Tse (✉) · J. Zhong · S. Fung
The Smart Engineering Asset Management Laboratory (SEAM), Department of Systems Engineering and Engineering Management (SEEM) City University of Hong Kong (CityU), Hong Kong, China
e-mail: mepwtse@cityu.edu.hk

J. Zhong
e-mail: jingzhong2-c@my.cityu.edu.hk

S. Fung
e-mail: samuelfung@karson-eng.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_39

between the significance level in implementing PAS 55 and the determining factors in asset information management section was explored.

Keywords Engineering asset management · PAS-55 standard · Information management

1 Introduction

Publicly Available Specification (PAS) 55, the new standard for asset management which developed by the Institute of Asset Management and the British Standards Institute(BSI). PAS 55 is becoming internationally accepted as the industry standard for quality asset management. The standard acts as a valuable guideline for asset lifecycle management, quality control, and compliance.

Good asset management requires meaningful, quality, timely asset and asset management information. Asset management information plays an important role for achieving an effective and efficient asset management system and for the continual improvement. It includes asset registers, drawings, contracts, licences, legal, regulatory and statutory documents, policies, standards, guidance notes, technical instructions, procedures, operating criteria, asset performance and condition data, tacit knowledge and all types of asset management records [1]. However, PAS-55 only lists a general guideline in what elements are required to be accomplished so that obtain the certification in EAM. There is no specific method to evaluate the standard implementations.

2 Background

2.1 Information Management in PAS-55

Asset management information is essential for achieving an effective and efficient asset management system and for the continual improvement of that system. The organization shall design, implement and maintain a system for managing asset management information. Employees and other stakeholders, including contracted service providers, shall have access to the information relevant to their asset management activities or responsibilities [2].

According to the standard, there are four procedures should be ensured in information management section. For application specification of PAS-55, they extended specific recommendations and guidance for each procedure. The questionnaire is designed due to the list of guidance and classified with four parts 53 questions. The detail of procedure showed in Table 1 [3].

Table 1 Main procedures in PAS-55 (Information Management section)

Part 1	Adequacy of information authorized for using
Part 2	Periodic review and revision to maintain adequacy of information
Part 3	Allocation of appropriate roles and responsibilities and authorities for information management
Part 4	Assurance of information security

2.2 Engineering Asset Management PAS-55 in Hong Kong

In Hong Kong PAS-55:2008 has been awarded the certificate by a number of public utilities corporations such as China Light & Power Co Ltd (CLP), Mass Transit Railway Corporation (MTRC) and China Town Gas Co Ltd (TG) etc. to cater for large scale investment and top demand of reliability for their EAM providing vital services to the Community. However in the general E&M building, services building and plants as a substantial part of EAM in the Community have not adopted the PAS-55 yet.

In this research case, a structured questionnaire was designed, and data were collected from 30 Operation and Maintenance (OM) departments for asset management in Hong Kong. It is to survey not only the real situation of asset management performing PAS 55 standard in information management part, but also analyze their effects and performance. The objects of investigation included: Public Services of HKSAR Government, Commercial, Residential, Industrial and Composite buildings portfolio in Hong Kong.

The structure of questionnaire is showing in following chart (Fig. 1).

All questionnaires were sent to 30 Operation and Maintenance (OM) departments, The objects of investigation included: Public Services of HKSAR Government, Commercial, Residential, Industrial and Composite buildings portfolio in Hong Kong. The data of survey are collected by Mr. Samuel Feng, who is the EngD student and he has almost 30 years working experiences on maintenance and asset management. The charts following showed the background information of respondents (Figs. 2, 3 and 4).

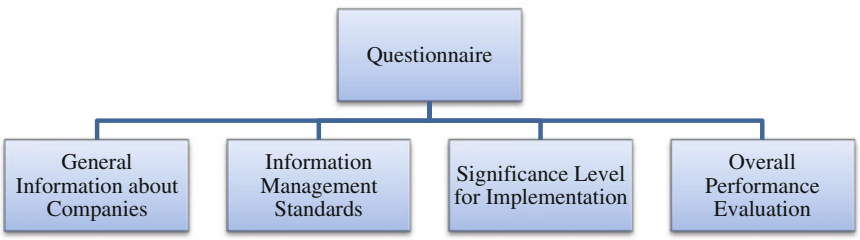


Fig. 1 Structure of questionnaire

Besides filling questionnaire, face-to-face interviews were conducted by the researcher, which could explain more about the purpose and the real meaning for respondents. It improved accuracy and efficiency. For another, some questions needed to be discussed for the real reasons indeed in order to obtain the information comprehensively.

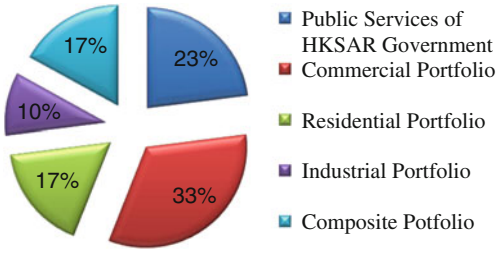


Fig. 2 Company categories of operations

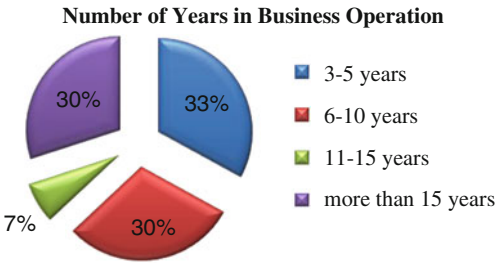


Fig. 3 Number of years in business operation

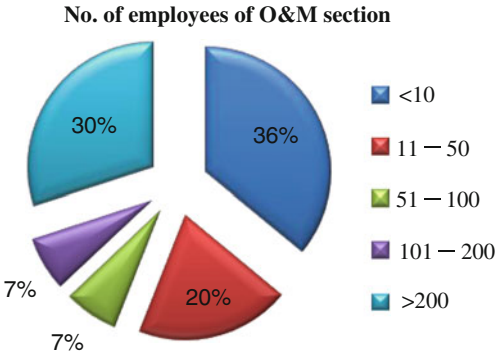


Fig. 4 Number of employees

3 Data Analysis

After the interviewer finished questionnaire, we have to check if anything missed or overlooked. During the whole process, interviewer communicated with researcher and ask explanation all the time when they were filling the questionnaire. Consequently, we consider the integrity and validity check was completed. To clarify the results of data, we defined and simplified four factors to evaluate the information management part according to the standard, which are:

(a) Adequacy
(b) Review and Revision
(c) Allocation
(d) Security

For each question, it indicated the degree to which the auditor agrees or disagrees with the statements, which is corresponding to one of the following 5 scales and 2 other categories, namely:

As mentioned above, there are 53 questions in total for measuring these four parts. For another, one section measured the influence for each factor in asset information management. It showed the significance of Information Management System of BSI PAS-55 in use for the O&M of the engineering asset (Table 2).

3.1 Current Situation

After questionnaire collection, the preliminary analysis was conducted, the current situation of all respondents performed in standard adoption is the following. According to the guideline mentioned above, there are four procedures which analyzed one by one.

The chart is the Adequacy distribution, most of them are between level 2 and level 3, which means the percentage of adoption is from 40–90 %, the highest is around 95 %, but the lowest is about 25 %.

For Periodic Review and Revision part, it is more decentralized compare with others. The performance of public service of government is lower relatively; however, the commercial sector has higher adoption level, which is around 85 %.

Table 2 Scale description

Totally adopted (100–91 %)	Mostly adopted (90–75 %)	Generally adopted (74–41 %)	Slightly adopted (40–11 %)	Not adopted (10–0 %)	More than those adopted	Neutral/ No need
1	2	3	4	5	6	7

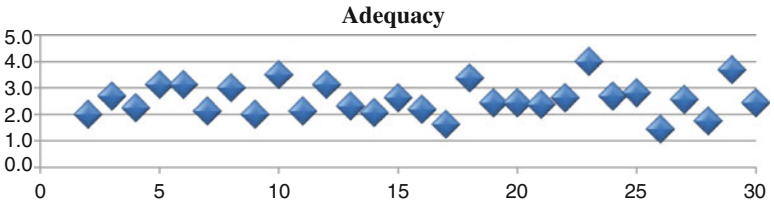


Fig. 5 Adequacy of information authorized for using

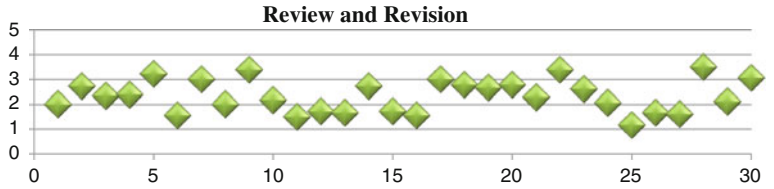


Fig. 6 Periodic review and revision to maintain adequacy of information

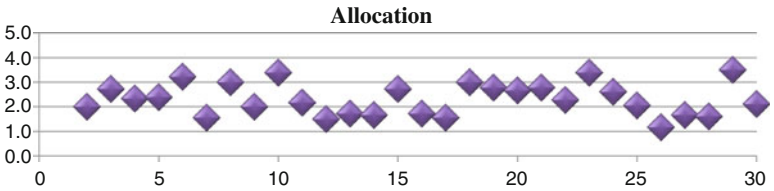


Fig. 7 Allocation of appropriate roles and responsibilities and authorities

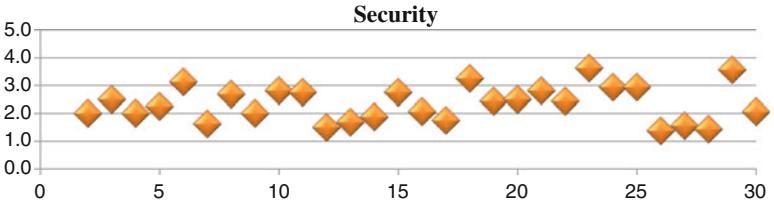


Fig. 8 Assurance of information security

The performance of Allocation part is better than others. Most of companies are in level 2 and level 3(55–85 %), the same as mentioned before, the best performance is banks.

In Security, they paid more attention on information security and the average score is higher than other factors, a majority of companies complied standard over 40 % (Figs. 5, 6, 7, 8).

From the charts above, they showed the real data in four aspects. Most of them are located from level 2 to level 3, which means the percentage of standards adoption is from 40 to 85 %. There is no response fully adopted or not adopted.

After data analysis, the bar chart expressed the commercial and industrial portfolio had better performance compared with others, especially bank buildings and data centres. The composite portfolio needs to improve their information asset management systems (Fig. 9).

3.2 Correlation Analysis

In order to check the influence of information management performance, Pearson's correlation analysis was conducted. Pearson's correlation analysis was used to measure the strength and direction of the linear relationship between a pair of quantitative variables. The correlation coefficient, calculated using SPSS, was used to determine whether there is any evidence of statistically significant association between these variables [4]. Pearson's correlation coefficients can take on values ranging from −1 to +1, −1 means perfect negative relationship, +1 refers to a perfect positive relationship. Positive correlation shows that as one variable increases, so too does the other. Negative correlation shows that as one variable increases, the other decreases. A correlation of 0 on the other hand indicates no relationship between the two variables; thus knowing the value of the first variable provides no assistance in predicting the value of the second variable [4].

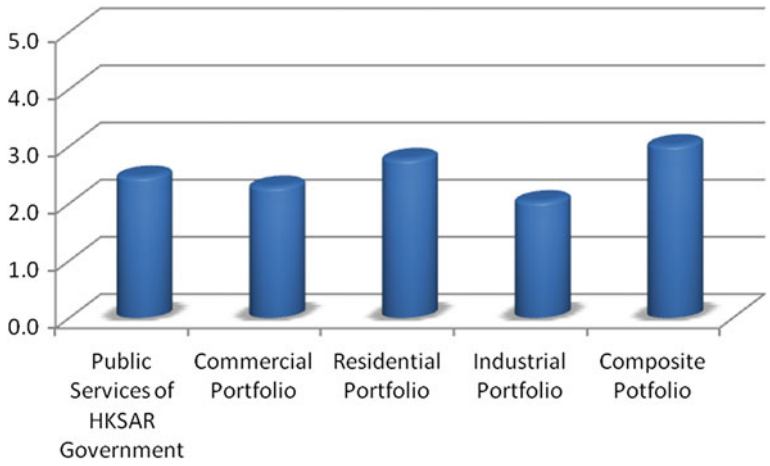


Fig. 9 Performance of different categories

In this study, the criticality of selected variables is decided by the significance of the correlation coefficients. A variable is considered critical when it is significantly correlated to the performance metric at $p < 0.05$.

The results indicated that there were four elements of information management in asset management; namely, Adequacy, Review and revision, Allocation and Security which are positively associated with Significance level (Table 3). It can be argued that these four elements focus on: Adequacy ($p < 0.05$), Review and revision ($p < 0.05$), Allocation ($p < 0.05$) and Security ($p < 0.05$), and are all directly involved in the Significance level of adopting PAS-55. Moreover, the findings also indicate that the most important practice that explains the variance in significance was Security (0.906) and Allocation (0.823) was significant at the 1 percent levels ($p < 0.01$). It showed information security is the most positive variable affected independent variable.

3.3 Gap Analysis and Performance Analysis

From the charts following, they showed the comparison between factors and their influence in asset information management. The diamond points are the real adoption for each company; the square points are the significance level of factor.

There are two scenarios:

The first one is that the gap existed in high adoption with low significance level. It means the companies or buildings have high percentage applied standard, but some parts are not critical for companies to adopt specifically. For this scenario, the reasonable solution is that reducing resources on some criteria and put more effort on critical processes in order to improve efficiency and effectiveness.

The second scenario is the gap found in low standard adoption but high significance level. This situation needs to pay more attention, which means the issues are important for companies, or it is high influence for no applying criteria. However, they did not follow standard in a good way, even they ignored some key

Table 3 Correlations analysis

Correlations						
		Significance	Adequacy	Review and revision	Allocation	Security
Pearson correlation	Significance	1.000	0.663	0.669	0.605	0.906
	Adequacy	0.663	1.000	0.787	0.762	0.748
	Review and revision	0.669	0.787	1.000	0.901	0.796
	Allocation	0.605	0.762	0.901	1.000	0.823
	Security	0.906	0.748	0.796	0.823	1.000

processes in asset management. The suggestion is that comply the standard specifically, and taking into account performance assessment timely.

Furthermore, we did ranking for their performance. Better Performance refers to high percentage of standard adoption and matching with the significance level, which means if this aspect is important to company, at the same time it has high adoption degree. On the contrary, some companies with lower adoption need to improve their performance to match their requirements, it is very critical for asset management and this is the main purposes of evaluating their performance (Figs. 10, 11, 12 and 13).

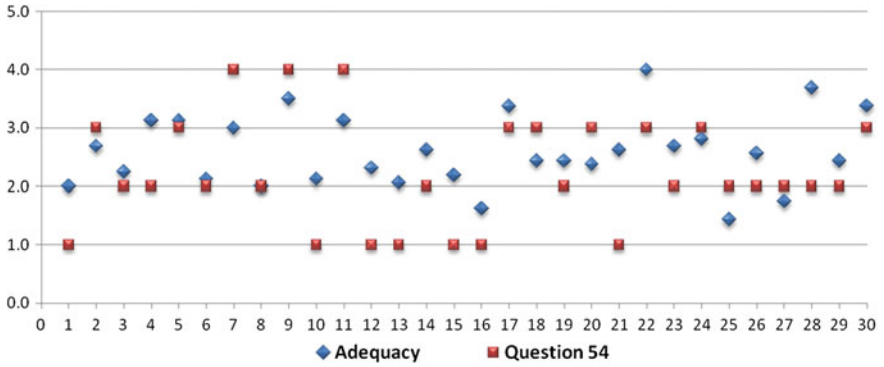


Fig. 10 Gap analysis of Adequacy part. Performance analysis in Adequacy part: Better performance (Company ID): 3, 6, 8, 16, 25, 27. Need to Improve (Company ID): 10, 12, 13, 15, 21

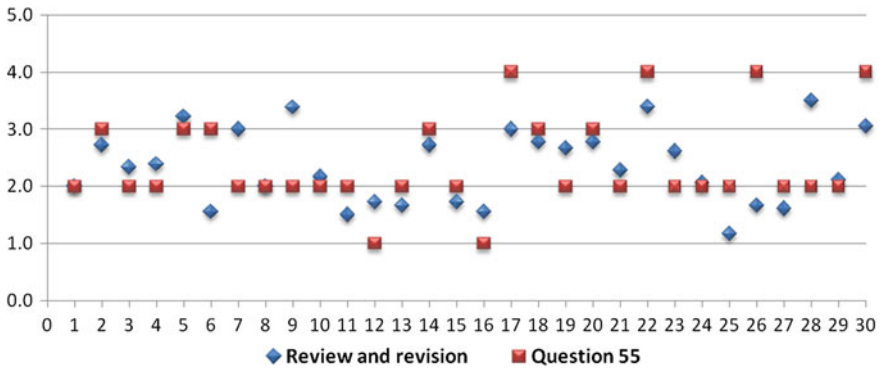


Fig. 11 Gap analysis of review and revision part. Performance analysis in review & revision part: better performance (Company ID): 1, 6, 8, 10, 11, 13, 25, 27. Need to Improve (Company ID): 12, 26

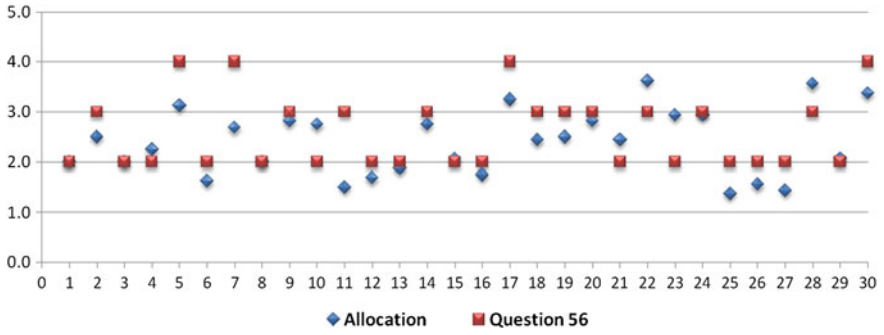


Fig. 12 Gap analysis of allocation part. Performance analysis in allocation part: better performance (Company ID): 1, 3, 6, 8, 13, 15, 25, 26, 27, 29. Need to improve (Company ID): 10, 23

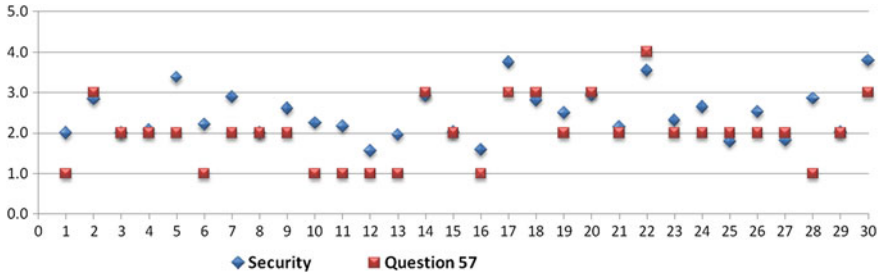


Fig. 13 Gap analysis of security part. Performance analysis in security part: better performance (Company ID): 3, 4, 8, 15, 25, 26, 27, 29. Need to improve (Company ID): 1, 6, 10, 11, 13, 28

4 Conclusion

This study investigated PAS-55 standard adoption in Hong Kong. The real situation revealed that most of companies, buildings or departments implemented the standards from 50 to 80 %. Furthermore, there is a gap existed between the standard adoption and the significance level. We performed gap analysis and ranked their performance. As a result, better performance refers to high percentage of standard adoption and closely matched with the significance level. The general suggestion for these respondents is to reduce resources on less important criteria and put more effort on critical processes in order to improve the effectiveness. On the contrary, some companies with lower adoption need to improve their performance to match with their requirements as it is very critical for asset management. In order to check the influence of asset information management performance, correlation analysis was conducted. The results indicate that there are four elements of information management in asset management, namely, the Adequacy, the Review and Revision, the Allocation and the Security which are positively associated with the

significance level. The results point out that the most important practice that explains the variance in significance is in the Security and the Allocation ($p < 0.01$). They also show that information security is the most positive variable that affected the independent variable.

Acknowledgments This article was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 122011) and a grant from City University of Hong Kong (Project No. 7008187).

References

1. Texas state auditor's office, methodology manual, rev. 5/95 data analysis: analyzing data—inferential statistics—1
2. British Standards Institute (2008) PAS 55: asset management. part 1: specification for the optimised management of physical infrastructure asset
3. British Standards Institute (2008) PAS 55: asset management. part 2: specification for the optimised management of physical infrastructure asset
4. Ling YY, Low SP (2009) Key project management practices affecting Singaporean firm's project performance in China. Int J Project Manag

Distributed Pre-processing of Telemetry for Mobile Engineering Objects

Vitalii Iakimkin, Aleksandr Kirillov and Sergey Kirillov

Abstract To solve the problems of prognosis and further cost optimal planning of strategy EAM the various approaches of analysis of objects are used. The two main directions for the analysis of the system of states are allocated. First is related to cloud processing of telemetry of engineering object. Usually, it is implied that to identify predictors of failure and determining their trends needs large computing resources. The second direction is characterized by necessity of urgent decision-making, so the system of diagnosis and prognosis should work in real time and should be adapted on-board. However, questions of reliability of prognosis, diagnosis arise at the use of recognizing automata in the form of neural networks, hidden Markov chains, and Bayesian networks. The ideal solution of the problem is reduced to the integration of peripheral automata and remote cloud computing in a single system. In this case, the remote PHM cluster besides the functions of computing long-term prognosis and the development of optimal cost-effectiveness maintenance strategies executes. Also functions of learning and retraining of recognizing on-board automata at change of the operating conditions of engineering or change of the physical state of engineering, do not increasing the risk of failures. Using a chronological database of individual objects, as well as a common database of similar objects, remote computer system analyzes the nature of the changed conditions or conditions of engineering and on the basis of CH&P model selects a system of evolution equations for prognosis of the development of predictors and hidden predictors. Then the method of retraining on-board recognizing automata is based on the evolution equations or the new scheme of real time recognizing that does not require training is constructed. The paper gives the experimental results demonstrating the operation of the automata and the remote cluster for commercial vehicles.

V. Iakimkin (✉) · A. Kirillov · S. Kirillov
Smart Sys Prognosis Center, Moscow, Russia
e-mail: SmartTechAppl@gmail.com

A. Kirillov
e-mail: SmartTechAppl@gmail.com

S. Kirillov
e-mail: SmartTechAppl@gmail.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_40

1 Introduction

Let's take as a well-known fact that the efficiency of engineering resource management and preventive maintenance are determined by the completeness of information about the functional state of engineering and the ability to accurate prognosis in the development of different functional states. In this case, the prognosis of functional states implies the following: based on data from on-board sensors of engineering determination of future permissible functional states in order to avoid the functional states of dysfunction, which reduce the engineering life and vice versa the determination of functional states which maximize the engineering life with the condition of preservation of its basic characteristics of work [1]. In the terminology of [2], the functional state is a trajectory of multi-dimensional segment of the wavelet coefficients from sensor of data obtained under conditions of continuous or periodic monitoring. Let's take also that the general purpose of monitoring systems is simplistically: asset effectiveness–the need to extract maximum profits from the minimum investment in plant and equipment. Referring to [3], and cited therein sources necessary to consider also the following factor, namely factor of the return on investment (ROI) for all kinds of monitoring. Briefly listed above conditions define the consumer interest in monitoring products and is an incentive to optimization of strategies of monitoring systems and requirements imposed on these systems. For today in asset of monitoring systems has: On-board sensors, including vibration sensor, pressure and rotation sensor, maybe also smart sensors, i.e. sensors with preprocessing of input signals. There is also a possibility of on-line mode when the sensor signals are transmitted to the remote server for further processing. A small on-board computing resource also exists, and the remote server has access to the service of cloud computing. In the listed conditions for monitoring systems the following problem is formulated: determination of the functional state of engineering (states), the prognosis of development of the sequence of states (determination of the trajectories) and management of trajectory, i.e. determination of the temporal dependences of the management parameters in order to select the most favorable trajectories optimizing the temporal characteristics of trajectories or maximizing the residence time of engineering on preassigned trajectory. Determination of the temporal dependences of the management parameters for self-maintenance thus should be cost effective. Cost effectiveness in this case means, that the total cost of data transmission, computing costs, additional sensors and measurement should be 10 percent of the economic effects, and time of return on investment should be a maximum of 5 percent of the life cycle of the insurance operations phase. Concrete calculations of economic efficiency and the time of return on investment in each case should be determined by taking into account a variety of additional conditions. Here the subject is designated for a correct formulation of the monitoring task. So, the next task is actualized: using the above possibilities in the form of sensors and various types of computing resources to optimize the task of monitoring, in other words, to minimize the total cost on monitoring under the condition of maximizing profits from reduced downtime, increasing the life cycle.

The present paper is devoted to describing the general architecture and algorithms for monitoring systems, basic recipes and conclusions made on the basis of the pilot version of PHM remote computing cluster. The purpose of the pilot was the following: on the basis of the above possibilities i.e., on-board sensors and on-board computing resource, possibilities on-line as a service of cloud computing to define the architecture and functionality of the monitoring system. The following tasks were set: to minimize the cost of supporting the monitoring system, the definition of long-term monitoring strategy aimed at identifying the most effective monitoring, scenarios, modes of monitoring, determination of the structure of databases, chronological databases of individual objects.

Chapter 2 is devoted to the description of computing models of the remote PHM cluster and algorithms. All of the described computational models are based on the principles of temporal hierarchy of a set of predictors of failures, dysfunctions, degradation of the material.

Chapter 3 is devoted to the definition and the necessary properties on-board recognizing automata, as well as methods for their learning and retraining.

Minimizing scheme of the optimal operation of the remote cluster and on-board recognizing automata is described. The main focus is on the development of this scheme on the market.

2 Remote Calculating PHM Cluster

The constructed architecture of computing cluster is shown in Fig. 1. At its core it is a traditional scheme of construction of remote computing, and the main interest represents the functionality of architecture because the further optimization concern namely areas of functionality understood here and below, as a sequence of methods and algorithms in the processing of distributed data between the cloud and on-board recognizing automata.

The main ideas, concepts of models for signal processing and the construction of their algorithms are defined by the authors [1]. The construction of processing

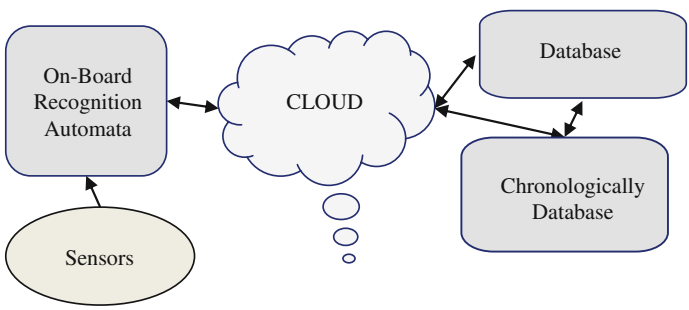


Fig. 1 Architecture of the monitoring system with PHM cluster

models in cited papers is based on preprocessing of the signal with its further processing by the hierarchical PHM algorithms. Preprocessing is the following sequence of actions:

1. Secondary discretization of signals with purpose of their unification and synchronization in each period;
2. Wavelet transformation
3. Determination of discrete processes (cascades) of wavelet coefficients with fixed the scaling and time (or angle) indices relative to the period number.
4. A set of histograms for the construction of dimensional and multidimensional empirical distribution functions for cascades.
5. Transition to vector processes determined by cascades segments of fixed length

The next steps of signal processing are based on the determination of the stochastic properties of cascades, and determined by them continuous stochastic processes.

It should be noted the two main points:

- Analysis of the cascades or processes represented by cascades;
- Analysis of vector processes.

It is implied that in the analysis of vector processes the empirical distribution functions of one-dimensional cascades (or processes) are invariant relative to displacement in time. In those cases where the given invariance is absent distribution function of one-dimensional processes, as well as vector processes are analyzed simultaneously. While the observed signal is non-stationary, constructed on wavelet coefficients with a fixed temporal and scaling indexes the cascades are usually stationary processes in those cases when all manageable engineering parameters are fixed and the external loads are stationary, unlike the case when there are transition processes. The condition of invariance of the one-dimensional processes (basic processes) determines the temporal hierarchy in the class of vector processes. Temporal hierarchy of vector processes determines the temporal hierarchy on the set of predictors characterizing the trajectories of vector processes, leading to accelerated degradation, dysfunction, and failures.

Thus, the temporal hierarchy of vector processes determines the temporal intervals of operation process of engineering, on which prognosis is not changed. The change of prognosis occurs when crossing the boundaries of hierarchical class by the trajectory of process. Crossing the boundaries of class by the trajectory is related to the change of stochastic characteristics of the process and, therefore for further prognosis the change of models and algorithms is required.

Prognosis models in each of the hierarchical classes are reduced to the definition of evolution equations for the empirical distribution functions, and in the case of its temporal invariant (the stationary) to the definition of the evolution equation for the density of the transition probability of vector states. The definition of evolution equations occurs by analyzing the stochastic properties of the processes and then determined the theoretical probability density or transition probabilities or their characteristics in the form of moments, cumulants, rational expressions of moments,

Fig. 2 (a, b) a—Set of interface windows reflecting the characteristics of basic and vector processes **b**—Interface windows reflecting text-graphic messages of logic automata

entropy characteristics, etc. Set of characteristics shown on Fig. 2a, is far from complete, and if added to it a list of the characteristics of vector processes, then abundance of interface windows is estimated to be several hundred. This number of the characteristics is impossible to analyze the operator in real conditions, so the processing algorithms have the sets of logic modules or Boolean automata, automatizing the analysis process. The second problem of modules includes the formation of text—graphic messages about identified predictors, messages about assessment of their trends and prognosis of development. The theoretical distribution functions or their characteristics are compared with the empirical distribution functions in different functional and probabilistic metrics. At an invariance of characteristics of the process the temporal dependences of the computed characteristics are known and thereby temporal (prognosis) dependences of characteristics on the time are known Fig. 2b. Some analytical expressions for the RUL assessment contained in the works [2, 3].

Defined above class of hierarchical models is the basis for construction of the basic functionality of PHM cluster and monitoring regimes.

Just following sequence of work of algorithms is the basis for further optimization of the monitoring and distribution of functionality between the PHM cluster and on-board recognizing automata, as shown on Fig. 3:

I. preprocessing described in the sequence (1–5)
II. algorithms of determining the types of processes for all wavelet cascades
III. comparison in chronological database and detection of stochastic processes with changed characteristics, i.e. empirical probability distribution functions of the basic processes
IV. determination of new evolution equations of probability distribution functions of the basic processes
V. determination of the temporal dependence of theoretical distribution functions, or its moments

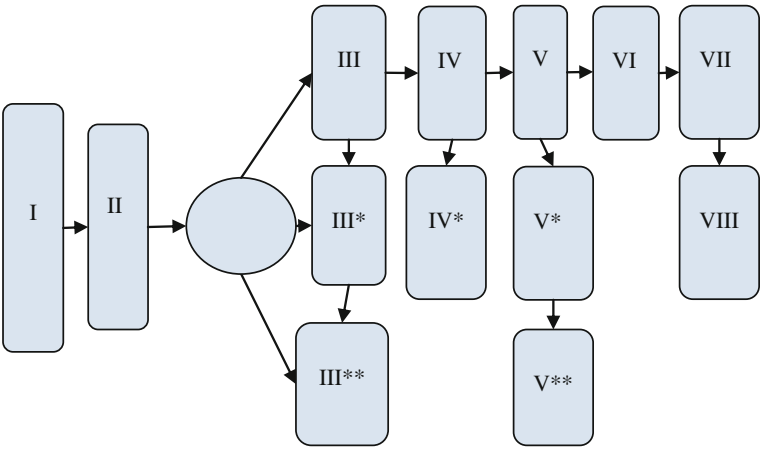


Fig. 3 The sequence of signal processing algorithms in the cloud

VI. comparison of theoretical and empirical results for the distribution of the previous time interval
VII. checking the preservation of properties of stochastic characteristics of process
VIII. calculation of the theoretical probability distribution functions, RUL to the boundary of the class (prognosis)

III* determination of evolution equations of the probability density of transition of vector processes;

Further calculations repeat IV–VII, but for transition probability

III** approximation of the stationary distribution functions by the exponents of polynomials

IV** calculation of the bifurcation sets;

V* determination of the close degree of the polynomial coefficients to the bifurcation sets;

V** Further monitoring of the polynomial coefficients.

Crossing the boundaries of class by the trajectory is related to the removal of degeneracy. That is, new predictors appear in a subsequent class, and they are qualitatively characterize other signs of dysfunctions, the degradation of the material, failures. Previous class does not contain such predictors because they are degenerate relative thereto. The consistent removal of degeneracy in the operation period is getting closer the system to the boundaries of failure, thereby reducing the temporal evaluation of the achievement of this boundary. That is, classes of states with a strong degeneration contain and earlier predictors of failures. Therefore for cost optimization of maintenance the earliest predictors are important to detect, and procedures of maintenance or self-maintenance conduct already at this stage. The transition of the system to less degenerated classes of trajectories changes the nature of self-maintenance and maintenance making it more high-cost. Thus, the results of

hierarchical models of signal processing allow to formalize the problem of optimizing maintenance and thanks such formalization to make them available for automation of the process of decision making in the choice of engineering operating strategies and in the choice of maintenance strategies for the entire life cycle of engineering.

However, target operating installation, change of external conditions, the need of engineering operation on the higher loads, sometimes reaching critical values, change and speed violations makes the problem of optimization of multi-valued and leads to the need for continuous or event monitoring throughout the life cycle. In addition, at the end of the operational period one more problem of optimization, related to the recycling technology, arises. It is also required rigorous calculations with the assessment of maintainability of engineering or its subsystems. In fact, the prognosis problem is decomposed into two scenarios.

$$P(w) = const\ \exp(-\frac{U(w, \lambda, \sigma)}{\sigma^2}). \quad (1)$$

In the first case, when the set of states is inside the class, and prognosis, as well as the choice of the optimal management strategy is based on the evolution equations. At the same time, the management problem is reduced to the determination of the temporal dependence of the management parameters in order to keep the system on the optimal trajectory. In the second case, at the crossing the boundaries of class, the nature of evolution equations changes, and with it the other and later predictors of failures are appear. At the continuous monitoring the time of crossing the boundaries of class is determined. However, this process of determination has costly nature, requiring continuous telemetry transfer to a remote server. In addition, there are factors requiring the monitoring system response in real time. These factors include sudden external impact on engineering, having random nature. Under the influence of these factors the sudden crossing the boundaries of class by the trajectory is possible. Such processes are fast nature (catastrophic). In view of the fact that they are initiated by external influences, they are unpredictable. Therefore at such events the rapid determination of the consequences and reassessment of prognosis or change of maintenance are required. The sudden change i. e. the sudden crossing the boundaries of class nevertheless is possible to analyze by methods of catastrophe theory [4]. There is an obvious benefit of such analysis for the following reason. Models of sudden changes in the density of the distribution function determines sets in the parameter space called the sets of bifurcations or catastrophes [5, 6] Fig. 4a. On the basis of the observed empirical distribution function can determine the degree of closeness of the management determine the degree of closeness of the management parameters and numerically evaluated external perturbations to the set of bifurcations. Thus, the values of permissible external influences and therefore changes in the degree of risk are estimated. On Fig. 4a, sudden changes of probability density with two parameters λ, σ are demonstrated. Parameter σ corresponds to the value of external influence on the mechanism, λ—management parameter. At excess of permissible external

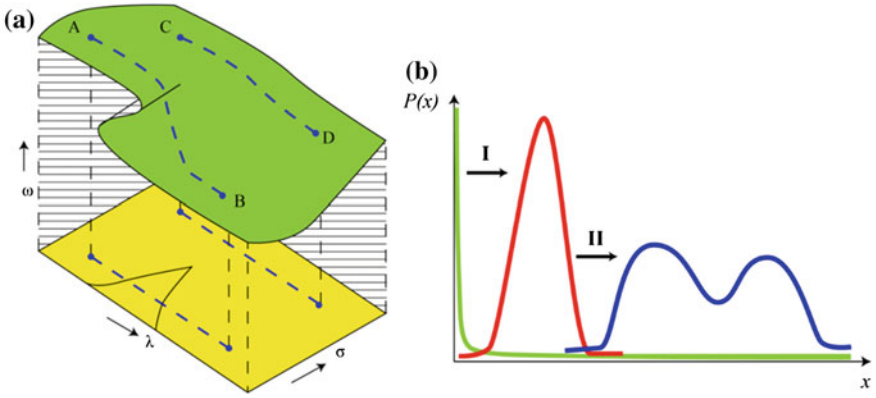


Fig. 4 **a** the sets of bifurcations or catastrophes, **b** the changing of type of the distribution function

influences the changing of type of the distribution function density occurs, as shown on Fig. 4b in the form of transition 1—loss of stability of the zero sign, at the transition 2—the appearance of bimodal distribution function. Itself bimodality means the generation and development of the failure, about what can testify the appearance of extraneous knocks in the engine, rotor, gearbox, etc.

3 On-Board Recognition Automata

Thus the process of continuous monitoring is necessary to optimize with attraction on-board computing resource. Placed on such resources recognizing automata can take over some functions of the remote computing cluster. However, in order to these computational functions have been implemented, it is necessary to construct the on-board computing and recognizing automata on the other principles. Traditionally understood role of recognizing automata by which is meant a neural network, hidden Markov chains, non-linear Wiener chains, Boolean automata, etc., are based on the recognizing of signs of failure of the observed signal.

In the problems of prognosis, it comes to the predictors of failures, i.e. signs that are appeared long before the failures themselves. In addition, the prognosis problems are formulated in high dimension spaces, that the problem of learning recognizing networks makes unsolvable, or requires compression of description up to dimensions in which the predictors are undefined. Partially recursively non-computable quantities, such as the Kolmogorov complexity is taken as predictors [3, 7]. Hence, recognizing automata are little use to solve the prognosis problems of the state of complex technical objects. This is evidenced by more than 20 years of experience of trying to create a diagnostic neural network for automotive engines

[8]. Besides, changes in external conditions, changes in the stochastic nature of external and internal noises with the necessity require retraining of automata. Meanwhile, attempts to create recognizing automata continues, but purposes have changed. From the purpose of creating a universal diagnostic neural network, the developers have put more narrowly focused tasks. Thus, in the situation examined here recognizing automata are aids. However, their presence in the on-board diagnostics promotes the appearance of multiple scenarios of remote monitoring with the transfer of part of the functions to the automata. This multi-scenario offers great opportunities to optimize and reduce the cost of expenses for remote monitoring. In addition, the above-described case of rapid change of states for engineering indicates the need of real time mode for the effective prognosis.

The automata must work not in the space of the observed signals, but on the set of predictors of classes. Therefore, the following objects lend themselves to recognizing.

1. Moments of change in the characteristics of stochastic processes, i.e. moments of crossing the boundaries of class by the trajectory including sudden, catastrophic crossing.
2. Recognizing of the trajectories within a class by describing their predictors. Here automata must be configured to recognizing the signs according to the principle YES-NO
3. Calculation of entropy characteristics of rare events.

But that's not all automata functions. In the process of monitoring and on the basis of processing of individual chronological database the set of predictors will be updated, the basic principles of retraining of can be formalized. Thus recognizing on-board automata are a family of automata with their adaptation under the target problem generated by on a remote cluster. The family includes automata, which recognizing the passage of the boundary, recognizing the trajectory with different entropic, capacitive and other characteristics, determining the change in length of segments and the vector processes, etc.

Different principles of learning and retraining are implemented in each automata of the family.

Neural networks: multi-layer neural networks are used to recognizing trajectories. In this case, the multi-layer neural network essentially is a locally-time approximation of the propagator or probability density of transition.

Hidden Markov chain is a good recognizing device for frequency analysis of rare occurring events, for the analysis of the trajectories without resorting to a set of statistics.

The bundle cellular automata implemented to recognize the evolution of failure regions during operation and to determine the bifurcation points of the trajectories in the problems of self-maintenance.

4 Conclusion

So, based on operating experience of the pilot PHM computing cluster results allow to do certain and practical conclusions about the architecture and the dynamics of monitoring systems intended for the prognosis of failure and being a basis for the development of self-maintenance systems, methods of preventive maintenance. The main conclusions of general nature are concern moments of the distribution of functions between the cloud and recognizing on-board automata. Discussed above distribution solves several problems of self-maintenance systems:

- Condition base maintenance
- Allows to formalized also prognosis models and basis for preventive maintenance
- For some perspective of self-maintenance

At the same time, such tasks as optimization of the maintenance and creation of low-cost on-line monitoring systems also become formalizable.

Division of functions to the extent learning with a gradual shifting on the shoulders of automata a growing number of functions gives the correct dynamics.

References

1. Kirillov S, Kirillov A, Kirillova O (2011) System of the automatic preventive on-line monitoring and diagnostics of car engines on the basis of the new methods of preventive diagnostics, sae world congress & exhibition. Detroit, USA, Technical Paper 2011-01-0747. doi:10.4271/2011-01-0747.
2. Khodos A, Kirillov A, Kirillov S, (2013), Multiresolution analysis time series data and RUL estimate, Chem Eng Trans, 33:337-342. doi: 10.3303/CET1333057
3. Kirillov A, Kirillov S, Kirillova O (2011) Algorithmic method of analysis of time series data for definition of prognostic parameters of engine fault. In: 3rd International conference on advanced computer control (ICACC 2011), pp 138–142. doi:10.1109/ICACC.2011.6016384
4. Kirillov S, Kirillov A, Kirillova O (2011) Theoretical models and market architecture of PHM monitoring systems. In: Prognostics and system health management conference 24–25 May 2011 (PHM-2011), Shenzhen, China, pp 1–8. doi:10.1109/PHM.2011.5939490
5. Arnold VI, Varchenko AN, Husein-Zade SN (1982) Singularities of differntiable mappings. classification of critical points, caustics and wave fronts. Nauka, Moscow, (in Russian)
6. Poston T, Stewart I (1998) Catastrophe: Theory and Its Applications. Dover, New York. ISBN 0-486-69271-X
7. Ray A (2004) Symbolic dynamic analysis of complex systems for anomaly detection. Signal Process, 84:1115–1130. doi:10.1016/j.sigpro.2004.03.011
8. Marko K et. al (1989) Ford motors company. Autom Control Syst Diag IJCNN
9. Kirillov A, Kirillov S, Pecht M (2012) The calculating PHM cluster: CH&P mathematical models and algorithms of early prognosis of failure. In: Conference on prognostics and system health management 23–25 May 2012 (PHM), Beijing, China, pp 1–11. doi:10.1109/PHM.2012.6228771

Sewer Linings—The Failures, Common Reasons and New Innovative Lining to Increase Reliability of Restoration

N. Subotsch

Abstract Concrete structures are an integral part of society. Today most sewerage systems are constructed from concrete as the foundation. Concrete is susceptible to corrosion under many conditions and in sewers the acidic waste can degrade the concrete asset quickly without protection, and lead to a failure of this foundation system. The Sewer Mains are predominant concerns however an often overlooked concern is with Man Holes, Wet Wells, Access and Inspection Chambers where the moist environment and stagnant air flow allows corrosion to readily occur. Considerable effort goes into the design and construction of concrete structures, the concrete pipes, sewerage pumping stations and sewerage treatment plants. These represent considerable financial investment and with clear appreciation of options to protect these assets and how they perform in the future will provide maximum operational life. There are various ways to protect concrete from corrosion in sewerage related installations. This is a simple overview of the common systems in use and outlines the various products strengths and weaknesses, and introduces a new novel approach based on traditional proven practices.

Keywords Man holes · Wet wells · Access and inspection chambers · Hydrogen sulphide · Sewer · Corrosion · Concrete · Waste water treatment plants (WWTP) · Rehabilitation

1 Introduction

Sewer mains, man holes, wet wells and other associated assets are all referred here to as underground asset. Concrete has been the predominant material of choice. It is relatively inexpensive, easy to cast or shape allowing it to be precast or laid in situ, and can be connected using a variety of techniques.

N. Subotsch (✉)
Peerless Industrial Systems, 2/79 Robinson Ave, 6104 Belmont, Australia
e-mail: nick@epigen.com.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_41

As a construction material, it is ideal but in underground assets, concrete can be degraded by corrosion by waste and effluent which often consists of a variety of chemicals including sewerage. To protect the concrete from deterioration, admixtures, linings, coatings, and other fixes are applied as a membrane or barrier.

This paper discusses primarily issues of existing assets and rehabilitation, the general conditions and fixes applied, the reasons for limited success and risks to long term service, and will conclude with an innovative means of increasing reliability in providing the protection.

2 Degradation of Concrete

Specifically the area of focus here is corrosion from internal flows, and does not include the corrosion that can arise from surrounding sulphate soils, or internally from AAR (Alkali Aggregate Reaction).

Corrosion resulting from internal flows can be expected to arise from chlorides and nitrates in trade or industrial waste, however sulphur is also arising from this waste. The most common or predominant cause encountered is sulphur from various organic and materials carried in our sewerage and is derived from the sulphates present, interacting with microbial materials. The havoc is often loosely referred as Sulphurous Acid but many different events and compounds can be present.

$$S_2^- + 2H^+ \rightarrow H_2S \tag{1}$$

Because under anaerobic (bacteria slime) conditions SO_4^- can form S^-.

The H2S gas rises to the vacant space and in the presence of oxidizing bacteria, have the H2S react with airborne O_2 to form H_2SO_4

$$H_2S + 2O_2 \rightarrow 2H_2SO_4 \tag{2}$$

H_2SO_4 can then react with the Calcium Hydroxide (Hydrated lime) resulting in Calcium Sulphate, being the loss of concrete structure.

$$H_2SO_4 + Ca(OH)_2 \rightarrow CaSO_4 + 2H_20 \tag{3}$$

H_2SO_4 is not in solution, but produced on the open damp areas of concrete and amount of acid produced (bacteria dependant) is in low concentrations and localized. The reactions take time; it is not a significantly large scale effect.

The following simplified diagram (Fig. 1) represents the activity in a pipe but is equally relevant in other assets.

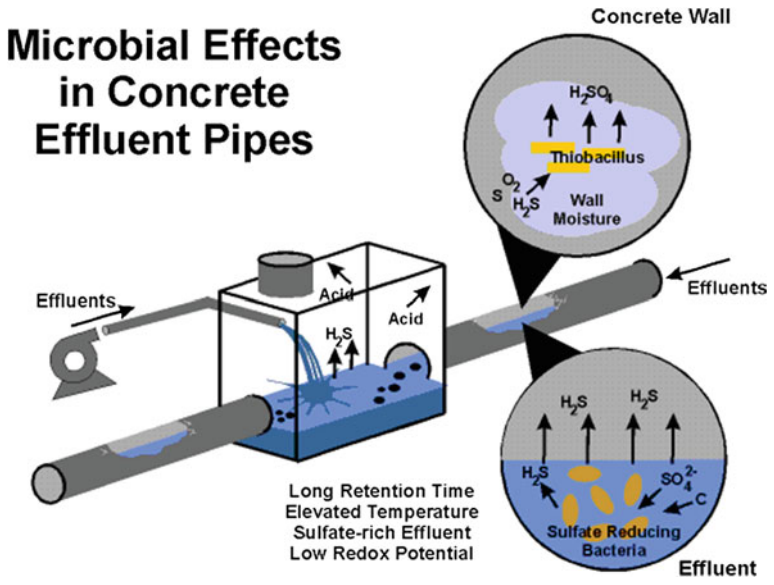


Fig. 1 Simplified diagram represents the activity in a pipe

3 Protecting Concrete from Degradation by Corrosion

Although various options existing in protecting concrete from corrosion, those relating to dosing the flows depend on consistent monitoring and accurate dosing to the corrosive compounds present. It is impractical unless batch treatment is possible and given the potential volume of sewerage involved an expensive effort.

Even treating the sewerage at a designated WWTP relies on coatings and linings when enclosed facilities or odour control is practiced. The concrete is best protected by a barrier.

There are many materials capable of doing this, thermoset and thermoplastic materials, rubbers, cement compounds, and a whole host of hybrid materials. Steels and their alloys are rarely used due to the inability to cover the range of chemicals potentially in the exposure stream, installation is difficult, and those that are suitable are very expensive.

Typically, the materials more commonly used are the following:

1. Epoxy—2 pack coatings and linings.
2. PVC—Polyvinyl Chloride preformed plastic.
3. PE & HDPE—Polyethylene and High Density Polyethylene.
4. Cements, Polymer Cements or Aluminates—Lime and related reactive cement bases

It is accepted there are many more varieties used in different parts of the world, and many hybrids of those mentioned but the greatest proportion are the 4 types

mentioned here. It would be difficult to do justice to them all, or this subject in the allotted time so I will not be referring to those.

4 Failures of the Common Materials

I refer to failures as the significant aspect of this paper because although each material has been found to be acceptable in providing protection in service, in sewerage rehabilitation it is the practical installation and ability for that material to sustain integrity over a broad range of conditions that limit life. Conditions that can cause premature failure is essentially the weak link and from which consideration made.

4.1 Epoxy

These are by far probably the most common materials in modern use of the non-sacrificial types that demonstrate both versatility and resistance to the exposures. Dating back to simple 2 pack paints, the difficulty in covering severely degraded surfaces resulted in failures from pinholes and loose substrate. These failures led to smoothing the substrate with sacrificial materials like cement render that altered the mode of failure to coating delamination from the render due to either contamination of the cement from the environment vapour during cure, the render itself letting go of the primary concrete due to poor adhesion, or the retained water in the render blowing off the lining. Epoxy based mortars are often incorporated to provide a smooth finish but the time to overcoating allows contamination of the surface and possible peeling. Subsequently, higher build epoxy were introduced which still had difficulty providing a proper barrier over extremely damaged concrete. Further developments led to fast cure epoxy types that allowed heavy builds over semi cured epoxy allowing the possibility of very high build but these systems are invariably exothermic in nature and the heat would draw moisture from within the concrete to the surface compromising adhesion. Alternatively the system itself on cooling would be contracting resulting in significant tension and the risk of failure referred to as mud cracking, or simply peel away from the existing substrate.

4.2 PVC

In many ways, these materials appeared ideal. Shaped or preformed and able to be welded, they could be slipped into place and back filled if necessary to bed into the surrounding asset. Over time, degradation of PVC through plasticizer migration leads to a more brittle material and the lack of adhesion to the substrate leads to

fractures that hide corrosive Hydrogen Sulphide ingress and degradation of base substrate, and thus inspection is never properly carried out. The longer term damaged can be extreme with collateral damaged to the asset or adjacent structures. The other issue with PVC is the inability to terminate properly the liner so the incidence of Hydrogen Sulphide vapour in behind the liner occurs often as condensation of the more corrosive acid resulting in a similar risk of asset damage and collateral damage.

There are other suggestions of PVC breaking down over time liberating Hydrochloric Acid and therefore itself being responsible for degradation of the asset, similar to that seen on steel roofing. This is a technical discussion with many varied opinions because those failures are centred on moist conditions and sunlight or heat. There is very little documented technical prof pertaining to this type of failure but it remains for the purposes of this paper technically feasible.

4.3 PE & HDPE

Polyethylene followed on from the PVC as a material that overcame greatly the longer term degradation and embrittlement issues of PVC. Over a brief period of time some failures caused by the PE creeping with thermal changes in the environment resulted in fractures of the welds, or large areas letting loose till fatigue led to ingress or Hydrogen Sulphide or sewerage inflow. These failures were quickly identified and the introduction of High Density Polyethylene quickly introduced. Many of these liners included tags or mushroom shaped clips that were formed or welded to the back of the HDPE and encapsulated by the back fill of mortar after installation, providing greater support across large areas. Instances of weld failure were still a potential risk, and ingress at the termination however, the advent of fabric backed HDPE which was welded to the existing liner allowed the fabric backed section of the liner to be terminated using epoxy adhesives and fibreglass.

There are now many types and variations of this type of liner being installed. In terms of resistance to the corrosive conditions in sewerage environments, they are marvellous. History is yet to decide the longer term suitability of the HDPE liners with many maintenance teams expressing the same misgivings as for any preformed liner, how do you check over time the ingress of inflow behind the liner and be able to address any seepage before substrate degradation and collateral damage on adjacent structures occurs. In simple terms, what you cannot measure you cannot manage.

4.4 Cements, Polymer Cements or Aluminates

Although these materials do not possess the chemical resistance to provide long term protection, they are finding renewed favour in rehabilitation work for the reasons of low cost. It is almost like the trends of change have come full circle.

Modern materials are giving way to the traditional cements and hybrids referred to as Polymer and Aluminates. The primary motivation is to rehabilitate the assets prior to loss of structural integrity and allow the lining to be sacrificial. This is not uncommon, we see the technical treatise of steel pipework in refineries where the corrosion rate of the pipes are measured and time to replacement planned. In a modern refinery these practices are as much a process requirement as it is a safety issue. The economics and operational parameters make it viable given the type of degradation experienced in a refinery whereas in the utilities domain, maintenance of live assets around a bustling community has limitations and a distinct demarcation exists between those that believe the new HDPE type liners can provide 50 years life and those that appreciate that nothing can go on without some monitoring and the costs associated with cement type rehabilitation allows monitoring and expense to be levelled as needed, where it is needed, containing the risk to the primary asset.

It is clear that each rehabilitation liner has its own advantages and disadvantages. The period of installation history for many is short in years so monitoring and reporting on the success of each is in the hands of personal choice and the technical debates influenced by marketing and financial concerns.

It is therefore realistic to postulate that 2 schools of thought exist. The first is to look at the materials that theoretically on paper provide 50 year life without consideration for the management of potential failures and collateral impact on adjacent structures. Will those that select these be around to concern themselves with any negative impact? The second is to revert to traditional liners whether they be fully resistant to the corrosion or be they sacrificial in nature, but importantly be monitored for maintenance but can necessitate an ongoing allowance for rehabilitation that will impact the community in relation to inconvenience and ongoing use of taxes and resources.

5 Innovative New Materials

The earlier reference to the trends of change having come full circle are not necessarily correct, and I expect trends will continue to be re-introduced time and time again. Cyclic yes, but at each cycle something more comes from the older material being reintroduced.

The times when epoxy was seen as the ideal material may be back among us with more value and advantage than ever before. Long life and the ability to monitor the integrity of any asset is a basic necessity, just like a medical check-up. The advent of renewed developments focused on better configuration provides the springboard to using technology that was fundamentally sound, and adapt fundamental installation practices to install linings with reduced risk of failure. Certainly more optimistic than new materials unproven in the field.

Personally, we are one of the organisations taking traditionally sound materials demonstrated that perform historically, and focused on how to place them easily and quickly.

The main challenges have essentially been:

1. Application to sound concrete substrate
2. Application fast enough to eliminate the effects of contaminants to the liner system
3. Application thick enough to provide adequate coverage that ensures a barrier free of imperfections
4. A material that has low exotherm so no tension or stresses are present
5. A material that can be quickly patched if ever required without total removal and replacement

Addressing each of these points more specifically.

1. The modern techniques of concrete preparation allow significant improvements in speed and thoroughness of preparation. The research work on concrete and being able to incorporate this into the procedure provides significant benefits. Expansive gypsum or ettringite can be overcome to arrest coating blowing off, and aggressive media can be added to high pressure water preparing the substrate to quickly remove sulphurous build-up to leave the concrete clean. Fans and dehumidification equipment can be employed to rapid dry surfaces.
2. Cement type liners can be placed in a very short space of time using modern equipment, so too can the coatings so intercoat exposure to the environment can be eliminated. Importantly the use of more conventional spray deposition equipment becomes more viable rather than heated lines and plural spray since the added complexity serves only to increase costs and reduce the number of people that can carry out the task without equipment failure or availability. Installing preformed plastic liners can be time consuming but the continuity of the liner as it is installed limits environmental influences and so they are suitable as materials under these criteria.
3. To date the materials that have been able to be applied at thickness to provide the coverage or integrity in laminate required to eliminate risk in continuity in the lining have been the PVC, PE, HDPE, Epoxy mortars or those loosely based on Cement type mortars. Epoxy required overcoating and it is the delay between coats that introduced adhesion problems and peeling. More recent Epoxy types can quite easily be applied at high thickness overcoming the shortcoming of Cements appreciating that the cements are sacrificial to the corrosion.
4. The latest and newest Epoxy based termed Ultra uses the same background systems as in the past, the proven performance experienced over the past 40 years has not altered so the recorded history over this same 40 years is well placed to provide the performance expectations in predicting asset reliability and life. The use of plural heated systems is not required alleviating exotherm, a critical factor in high thickness application because any system that cures hot

Table 1 Comparison of different coating systems

	Ease of application	Effects of corrosive conditions	High thickness	Low internal stress	Ability to patch
Traditional Epoxy	X	X		X	X
PVC		X	X	X	
PE & HDPE		X	X	X	
Cement based Mortars	X		X	X	X
Modern epoxy Type (Ultra)	X	X	X	X	X

must cool and therefore the post cure dimensional cross section will be under contraction load or tension - without balance.

5. Inevitably, any lining must undergo some form of damage whether it is a retrofitted ladder, screen, or fatigue. The Cement type liners and epoxy based have demonstrated that simple patching or repair can be carried out without a great deal of effort and without skills and equipment that make such an exercise beyond the scope of the average handyman. Welding of plastics in repair is possible but the skills of the workforce require training and the level of preparation is not always simple to get good adhesion or reduce weld embrittlement.

Reviewing the table (Table 1), only the newest Ultra material demonstrates the versatility across the important criteria listed. It is a subjective review but each point has been considered by reference to industry (refer Acknowledgements). The Cement based Mortars deserves special mention because of ease of inspection and maintenance but the new type of Epoxy Ultra appears to have a significant advantage. If the long term resistance to sewerage environment is considered, the new Epoxy Ultra is certainly a material that deserves greater consideration.

6 Conclusions

There are many materials as either hybrid types of the aforementioned, or consisting of entirely different technology that have been used in the rehabilitation of sewers. These have, from time to time included polyurethanes and polyureas, fibreglass laminates and methacrylates. Unfortunately they have not taken to the industry well and suffer from either confined space issues pertaining to flammable solvents or require primers or are water sensitive. Failures and difficulty with installation make them unsuitable for use across a broad range of practical considerations. We still see today the majority of discussion and proposals based around the same few materials that have been in use for decades, and will continue to be used for decades to come.

The newer preformed plastics will continue to be offered to the market and appear to have a place during new construction however ongoing issues pertaining to termination of the plastic and how to ensure any seepage behind the liner is still a concern in being able to identify this before major damage occurs.

The new Ultra technology has none of the shortcomings of earlier systems and has the demonstrated chemical resistance to be a firm advantage, not relying on the promise of 50 years maintenance free life—yet quite possibly achievable, providing the integrity of the asset itself is not compromised. Given the fact it lends itself to be able to be repaired in the field with relative ease leaves it a very practical consideration.

Acknowledgments Robert Callant, Melbourne Water, Vic Aust, John Boyle, Kempsey Shire Council, NSW Aust ,Daniel Leeming, PIS, WA Aust ,Graham Thomson, Barwon Water Vic, Aust, Roy Orr, PIS, Qld Aust, Bill Allen, Fabfit, Qld Aust , Michael Arnott, McElligott & Partners, Vic Aust

References

1. G Thomson (Barwon Water) (2000) Corrosion & Rehabilitation of Concrete Access Chambers (WIOA)
2. M Tullman (2010) Corrosion Club (www.corrosion-club.com/)
3. Lafarge Fondu International (2004) Lafarge Studies On Corrosion Principles
4. Baker CA (1974) Chemical Corrosion of Concrete. Australian Corrosion Association Conference, Hobart
5. R Orr, AGRU & STEULER, Technical Officer & Advisor, (2000–2013)

Bottleneck Management in Supply Networks: Lessons to Learn from a Synoptic Systems Perspective

Jakob E. Beer

Abstract Networks of organizations are the predominant form of value creation in manufacturing industries and are becoming more complex as complexity of products and specialization of knowledge increase. A higher level of collaboration among organizations changes the distribution and types of benefits and risks in supply networks. Some of the risks result from increased dependence on other organizations in the network as these take over larger portions of value creation or functions closer to the very core of the product. With a great share of value-added being created by suppliers, reliability of the supply network becomes a more pressing issue. Bottlenecks in such networks can turn into major impediments to the success of the focal firm. Drawing on literature on bottleneck analysis, production planning, systems theory and supply chain risk management, this paper examines some commonalities and differences between local production systems and supply networks with respect to the emergence of bottlenecks and how they can be managed.

1 Introduction

Several industries are characterized by high fixed cost due to expensive production assets and complex organizational structure. In such industries, shortage of raw material as input for production results in high idling cost. Firms put much effort into prevention of production halts. Transportation, for instance, is often expedited through change to faster transportation mode; in the automotive industry, for instance, it is not uncommon to have scarce parts delivered by helicopters to avoid production halt in OEM facilities. High idling costs of automobile production plants justify the enormous transportation costs incurred by air delivery. Generally, disruptions and delay in supply networks can be a strong impediment to financial success and their management is considered a major competitive factor [1].

J.E. Beer (✉)
Centre for Industrial Asset Management, University of Stavanger, Stavanger, Norway
e-mail: jakob.e.beer@uis.no

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_42

While there are proven methodologies for improvement of material flow and bottleneck management in production facilities, it seems that no standard body of literature has evolved to tackle similar problems of material flow in supply networks. In this paper, I adopt a bottleneck-centred perspective and compare two types of material flow systems—factories and supply networks—across several relevant system properties with respect to problems of material flow management. The objective is to demonstrate potential shortcomings as well as opportunities for improvement and to provide management with a clear vision on pressing issues in supply network planning.

2 Bottleneck Definition and Bottleneck Management

Although the literature on bottleneck is extensive, it turns out to be difficult to find a good definition of what a bottleneck is. The term is widely used in a variety of scientific disciplines, with production planning and control probably being the most relevant branch of research in the context of this paper. Many authors do not even provide a definition, but those who do often refer to a "machine" [2], a "physical hindrance" [3], a "process" [4] or a "process step" [5] limiting system throughput, or more abstract "any capacity equal or less than the demand placed upon it" [6]. I am proposing a more systems-oriented bottleneck definition:

> The bottleneck of a system is the element (node or graph) that limits the system in attaining higher throughput beyond a certain threshold. This threshold is determined by the bottleneck's physical throughput capacity, organizational rules, or operational practices.

Figure 1 illustrates the concept.

It is important to point out that every system has a bottleneck somewhere, otherwise its throughout became unlimited. Bottleneck management comprises all activities for the prevention, identification, exploitation, location, and elimination of bottlenecks. There are many different types of bottlenecks and the discussion of the differences is beyond the scope of this paper. For now, it will be sufficient to point

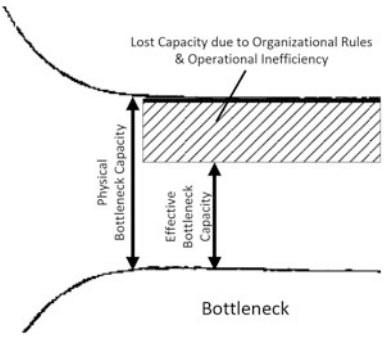


Fig. 1 Illustration of a Bottleneck

out that there are *planned* and *unplanned* bottlenecks and that the broad tasks of bottleneck management depend on the type of bottleneck we are dealing with. For unplanned bottlenecks (those that *emerge* somewhere in our system without us wanting them there) the important tasks are in prevention, identification, exploitation, and elimination, whereas for planned bottlenecks we have to think about exploitation and location (i.e., where to purposefully place a bottleneck in the system). The remainder of the paper deals with both types of bottlenecks as it will become clear from the context.

3 The Flow Principle in Factories and Supply Networks

Manufacturing has gone through paradigm changes in the past several decades. Much of this is owed to the convincing success of firms which adopted a different mindset towards production and employed production and management philosophies such as Lean, Total Quality Management (TQM), or Theory of Constraints (ToC). Production and management philosophies such as Lean and ToC have different methodological foci, yet they do share certain themes [7, 8]. One is the improvement of flow.

Material flow in a factory system can be understood as the transformation of inventory into throughput [8]. Improved throughput (i.e., improved flow) brings an organization closer to its goal of making money [6]. Continuous (one-piece) flow is the ideal situation Lean aims to create through the elimination of waste (muda) [9]. Accepting the focus on throughput and flow has been (still is!) a challenge for many production firms as it conflicts with the wide-spread acceptance of machine efficiency and cost reduction as primary objectives; nonetheless, philosophies such as Lean, TQM, and ToC have had considerable impact and the flow principle has been widely embraced.

Looking at the Supply Chain Management (SCM) discipline, one might easily get distracted by voluminous definitions putting forth claims about increased well-being of all the partnering companies and how everybody would win if they just did not only optimize for themselves… Such tales (which surprisingly often have become uncritically accepted text book claims) attempt to address organizations' credulity instead of their rationale. If we strip away all the normative claims of SCM definitions, we will find that SCM is essentially about the same thing as production management in a factory: to keep the material flowing.

4 A Systems Perspective on Factories and Supply Networks

Both factories and supply networks can be understood as systems of material flow. That is, factories and supply networks share certain characteristics. Yet, surprisingly little attention is drawn to the commonalities and differences between these two

types of systems with respect to bottleneck management and the optimization of material flow.

There are open systems and closed systems. Open systems allow material or energy to enter and to leave and thus allow change of system constituents [10, 11]. Both factories and supply networks show characteristics of open systems. Material flows into the system as input, experiences transformation, and subsequently leaves the system as output.

In the remainder of the paper, I will first describe some systemic differences between factory systems and supply networks before I will outline the implications of these differences for bottleneck management in supply networks.

4.1 Degree of System Openness

As mentioned in the preceding paragraph, both factory systems and supply networks can be conceived of as open systems. That is, both are subject to influx as well as outflow of material. Moreover, elements of either system can be affected by several kinds of impact from the systems' environment.

By looking at the spatial distribution of these systems' constituents (their nodes and graphs), we notice, however, that there are significant differences in the degree to which either system is exposed to possibly interruptive impact from the outside. On the one hand, we have often globally dispersed supply networks within which all nodes are subject to different environmental, political, economic, and cultural conditions with often long-distance transportation between several tiers; on the other hand, we have a series of assembly lines and machining stations in close proximity. This dichotomy is, of course, highly stylized, yet it makes certain differences easier comprehensible. Also, this does not mean that I suggest suppliers should be kept in high spatial proximity: Craighead et al. [1] found that geographic accumulation of suppliers is perceived as a risk by supply managers of large OEMs. The reasoning is that external effects that impact on geographic areas would affect all suppliers at once, incurring possibly devastating economic problems for the supply network. Examples are natural disasters, such as the 2011 earthquake and the tsunami off the northeastern coast of Japan, and political events involving larger geographic areas, such as the political turmoil in several states in North Africa ("Arab Spring") beginning in 2010. Thus, the risk of bottlenecks in supply networks due to the degree of system openness and spatial proximity in general and external economical, natural, and political shocks in particular could be visualized with a U-shaped graph.

Whereas risk of geographically concentrated supplier networks is somewhat obvious and systemic, risk of widely dispersed supplier networks is not quite as straightforward to grasp given the various sources of uncertainty. The more it seems important to support the planning process of widely dispersed supplier networks analytically so that the most important factors are taken into account. A comprehensive understanding of such factors may easily alter sourcing decisions as it turns

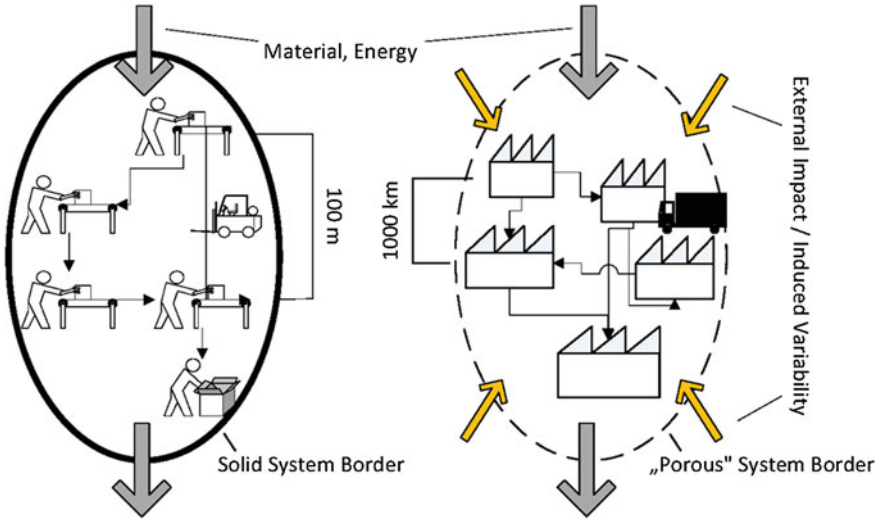


Fig. 2 System comparison: factory systems tend to have clearly defined system borders and higher spatial proximity between nodes as compared to supply network systems

out that transaction cost incurred by certain sourcing decisions can exceed anticipated savings.

Generally, while both systems are open systems, we can conclude that supply network systems are more open to their environment than factory systems and thus more prone to be subject to external impact. In other words, the system border of supply networks is "porous" (cf. Fig. 2).

4.2 Autonomy of Nodes

One important difference between management of a factory and management of a supply network lies in the autonomy of the nodes that are to be managed. Several authors raise the question to what extend networks can be managed at all and to what extend the focal firm simply has to cope [12, 13]. Obviously though, management of the focal firm will have to deal with a *lower degree of control of nodes in the supply network* than of nodes within its own organizational (and legal) boundaries. While a production firm can control its assets located in a local factory as well as its processes, it does not normally have the same amount of influence on suppliers' production assets and processes. It certainly is too simplified a notion that management has full control of the production assets the firm owns and the processes it deploys—much management research deals with opportunism and principals' attempt to control agents (e.g., workers)—yet I think it is a valid claim that in most firms management and owners will have considerable control of their assets and processes and thus can change things and can make change last.

On the other hand, OEMs often do not have *no* control over suppliers, either. In supply networks with OEMs that have considerable negotiation power due to their size and importance as a customer, OEMs do, in fact, intervene sometimes and actively change processes of suppliers. Also, OEMs can influence suppliers through the criteria they use for sourcing decisions. Supplier audits are another measure through which OEMs steer suppliers to their favored direction.

The process of interorganizational coordination and different constellations of power is often referred to as *network governance*. Network governance has been approached from different angles (cf. [14–16]). The different perspectives on network governance have in common that power configuration in networks can be diverse with OEMs being able to control and change how the network functions when they are able to control information flow and communication while suppliers do not possess resources (in the Resource Based View sense; cf. [17]) that would allow them to occupy a power position.

The higher autonomy of nodes comes with *limited access to important information*. The focal firm is dependent on the information its suppliers provide and it has only limited options for verification. Supplier audits are one attempt to gain "better" information, yet audits provide only temporary access and do not provide an adequate tool for day-to-day business. While in most cases suppliers certainly want to support their customers with the information they need, they might have strategic interest in not revealing some important information in other cases. An earlier study of the automotive industry [18] provides an example: Suppliers which were producing at maximum capacity due to high simultaneous demand from several customers saw themselves confronted with the problem of prioritizing customers. A supply chain manager from one OEM became suspicious because a competitor would receive parts which they were missing. Hence, he accused the supplier of yielding to the (larger) competitor's intense pressure. It seems unlikely that in such situation all relevant information (e.g., production capacity available and how it will be allocated) is openly communicated to all customers.

Additionally, OEMs and their suppliers as legally independent entities may encounter conflicting interests. The most obvious conflict is that OEMs tend to demand the lowest prices possible from their suppliers whereas suppliers need to maintain a margin to stay profitable. In terms of production planning, suppliers are confronted with two contradicting requirements: to produce as efficiently as possible so that cost per part is low and to maintain enough flexibility to never let the customer run out of supply and cause interruption of production. If we reframe this in terms of bottleneck management and again use the factory as comparison, the conflict of interest will become apparent: In factories, we may observe that production planners consciously design a bottleneck into the system. Often, they would choose the most expensive asset to be the bottleneck for reasons of depreciation. In a supply network where there is no almighty production planner with full control over all assets, which node in the network would become the bottleneck? By tendency, each individual firm would like to be the bottleneck (certainly without being an impediment to material flow in the network) and have high utilization to ensure "efficient" operations. From a material flow perspective, however, it is

beneficial to have buffers in the system that do not aim to be as tight as possible. This is not going to happen, however, in a supply network without centralized power.

4.3 Degree of System Complexity

Comparing complexity of two generic systems which have not been defined in detail is somewhat tedious. Of course, it depends on the concrete types of supply network and factory system we want to compare. There is a danger of tautological arguments of the type "Complex supply networks are more complex than (simple) factory systems", or vice versa. Hence, we either have to be very specific about what type of factory system and what type of supply network we are going to compare—or we confine ourselves to discussing factors contributing to complexity of supply networks which at the same time do not exist in local factory systems, which is the path I am choosing here. Due to page count limitations I will limit the list to very few but important.

The probably most important factor for a high level of complexity of supply networks has already been discussed: the degree of system openness. Supply networks simply are more prone to be reactive to external impact than a local factory system which is (literally) shielded by brick walls. Another important factor is the autonomy of the nodes as discussed in the previous section: firms representing nodes in a supply network system normally are subject to much higher autonomy than machining stations are in a factory (which are subject to fiat and control by management). The autonomy of the nodes gives rise to the complexity of supply networks as each node reacts to feedback given by other nodes (and actors external to the system) and at the same time provides feedback to others.

While both factory systems and supply network systems are material flow systems, obviously they do differ in scale; in fact, factory systems are sub-systems of supply networks. Spatial proximity between nodes in a factory is much higher than between two factories. Higher geographical distance, in turn, means more time is required for transportation of physical goods along the graphs connecting the nodes. Due to this delay in physical transportation there is an information gap. Modern interorganizational information systems attempt to close the gap. Nevertheless, there may always be surprises when goods arrive (from wrong parts to missing parts to defective parts…) because even the most advanced information technology can't prevent human error completely. Information gaps give further rise to complexity. One of the most prominent examples is the bullwhip effect [19].

Figure 2 illustrates the differences in terms of system openness and spatial proximity.

4.4 Focus on Flow

Significant improvements have been possible in factory systems as a consequence of a paradigm shift from focus on efficiency/cost to improvement of flow. The flow principle has been operationalized through measures such as the reduction of inventory—particularly of work in progress (WIP)—, the reduction of waste, and deliberate exploitation of bottlenecks. It seems, however, that in the larger supply network system the flow principle has hardly been implemented to a similar extent as in the smaller factory system.

In many supply networks, OEM have their suppliers deliver parts just-in-time (JiT) or even just-in-sequence (JiS)—which seems like the logical extension of the factory-internal flow system to the supply system. At second sight, however, it becomes clear that the just-in-time delivery adopted in many supply relationships is just this: *delivery*, while production on the supplier site remains only loosely coupled to customers' production and end customers' demand. Consequently, the well-known problems which Lean and ToC aim to resolve can still occur on suppliers' production sites and interrupt the supply network. It will not take the reader surprise that the positive effects from a change in delivery mode to JiT or JiS for suppliers are often small and sometimes even negative whereas OEMs often do greatly benefit [20]—an obvious contradiction to the supply chain management mantra of the benefits for all parties in the chain.

Another important aspect of material flow planning is the purposeful location of the bottleneck. Bottlenecks do not necessarily emerge at random places; stations may rather be designed as internal bottlenecks right from beginning of the material flow planning process. The advantage of such conscious design decisions is that throughput of the bottleneck (and thus the system) can be deliberately protected through buffers and machines with higher capacity [6, 4]. There are different approaches as to where the bottleneck should be placed and there are different arguments in favor of and against these approaches (cf. [21]). It seems, however, that supply network planning has not embraced the idea of purposefully designed bottlenecks yet; first results of an ongoing multiple-case study with manufacturing companies indicate that while the concept is understood upon explanation, interview partners (from purchasing and SCM functions) had never heard of it before. Following up this concept may open up interesting research venues.

Besides power-based control and fiat, OEMs can attempt to influence suppliers through cooperative behavior. If OEMs pursue improved material flow across organizational borders in a serious way, i.e. they try to get their suppliers not only to deliver their parts JiT but also to produce JiT, buffers between OEM and suppliers can be reduced which, in turn, forces the companies to closer coordination and stronger ties. Strong ties, in turn, may enable firms to pursue options in their relationships which would be outside their reach if they maintained an ad hoc arm-length relationship and thus would arguably provide even more options to improve material flow.

5 Implications for Bottleneck Management

The differences outlined above have implications for bottleneck management. The implications concern all the five aspects of bottleneck management: prevention, identification (detection), exploitation, elimination, and location.

Detecting bottlenecks is not a trivial task in either system. There are several methods for bottleneck detection. The reasons why there are several methods are three: (1) there are different definitions as to what constitutes a bottleneck, (2) there are different limitations in different systems which may render certain methods inapplicable, and (3) different methods may identify different nodes as bottlenecks and we do not know in advance which bottleneck corresponds to our understanding (our definition) of bottlenecks so we need different methods to double-check our suspicion and to avoid both false-positive and false-negative errors. Well-known methods for bottleneck detection in factory systems are utilization-based methods, queue length-based methods, wait time-based methods, and experiments [22, 23]. Literature on bottleneck detection in factory systems is voluminous.

Regarding the ambiguity of bottleneck detection encountered in factory systems we can expect even greater difficulties in supply networks for a variety of reasons. From the discussion of the autonomy of nodes (see Sect. 4.2) it becomes clear that from OEM perspective suppliers often are black boxes. Even if suppliers behave cooperatively, customers will still have to deal with incomplete information. Suppliers may even have strategic interest not to reveal certain information (and they have the autonomy required to protect their interests). Utilization-based methods, for instance, thus become difficult to execute in practice as the focal firm will not be able to get exact information about utilization levels of suppliers (as illustrated by the example from the automotive industry). By the same token, experiments may not be feasible to detect bottlenecks or to validate bottlenecks detected by some method.

Greater system openness, and hence higher complexity, may easily obscure the true causes for material shortages and make bottleneck detection even more difficult. In combination with lack of control over suppliers and with incomplete information the difficulties for effective detection of bottlenecks become apparent. The limitations for bottleneck detection in supply networks are manifold as compared to factory systems.

When a bottleneck has been identified, the objective should be to protect throughput of the bottleneck since throughput lost on a bottleneck is throughput lost for the entire system [6]. The objectives resulting from this insight are straightforward: don't ever let the bottleneck starve and don't waste the bottleneck's capacity. For factory systems, the drum-buffer-rope concept of ToC [6, 21] suggests having a buffer right after the bottleneck so the bottleneck will never be blocked, having stations of higher capacity right before the bottleneck so even in case of disruptions they can make up delays due to their higher capacity so the bottleneck will not starve, and releasing new material into the system only at the rate the bottleneck is working. Moreover, not to waste bottleneck capacity the bottleneck shall never work on defective parts (quality control should be placed right before

the bottleneck), it shall not destroy parts (diligent maintenance of the bottleneck's tools and work process is of great importance), and the parts it has processed shall be treated carefully in subsequent steps so they don't break. Also, it shall not work on parts that could be processed elsewhere nor shall it work on parts now for which no real customer demand exists or which lead to delays for other parts more urgently needed according to the due date. In order to implement such measures, focal firms would need to be able to exert significant power over suppliers, and not only over direct suppliers but over suppliers of higher tiers as well. In other words, focal firms would need full-fledged tier-n management with a very high level of control over suppliers of all tiers. Realistic? No. Not even the largest automotive OEMs which occupy power positions in their networks are even attempting it. Only very recently, a few OEMs have approached tier-n management [18]—yet certainly not to the extent required to achieve changes as profound as those outlined here.

Additionally, the high level of system openness and complexity in supply networks make it impossible to shield bottlenecks—or nodes that may turn into bottlenecks—from external, possibly adverse impact. Natural disasters, political uprise, and economic swings are just a few broad sources of disruptions that will occur in the widely dispersed supply networks we are often dealing with today.

As for bottleneck prevention, the situation is difficult due to higher variability and complexity in combination with less authority over entities and lack of complete information. Nevertheless, it seems that through sensible supply network design—preparation of alternative, flexible sources, creation of trustful relationships with suppliers, supplier auditing and training, match of product properties, process characteristics, and delivery mode, early communication of forecasting and sales data, to name just a few measures—the sudden emergence of bottlenecks can be prevented to a good extent. Hence, although a supply network tends to be more difficult to manage than a local production environment, a strong emphasis on bottleneck prevention in the supply network design phase is likely to mitigate the problem.

6 Conclusion

Some key points can be derived from the preceding discussion.

Supply networks show a greater degree of system openness than factory systems, i.e., they are susceptible to a greater variety of external factors. A higher level of external impact can, in turn, increase variability (or more general: uncertainty). Moreover, the basic units of supply networks—firms—normally show greater autonomy in their actions than machining stations in a factory while maintaining a lower level of transparency. The factors combined contribute to a higher level of complexity. Higher complexity, then, may obscure causes for the emergence of bottlenecks and may make bottlenecks more difficult to find due to a high level of "noise", i.e., other problems that occupy management's limited attention span. Identifying bottlenecks in factories is far from being trivial and it is likely to be

even more challenging in a supply network with more incomplete information, greater autonomy of actors, and arguably high levels of opportunism. Likewise, firms are less able to influence and exert control over other firms' processes and assets in supply networks than management in an individual factory is able to influence and exert control over local processes and assets. These factors vary in intensity case by case, though.

The focus on flow has been identified as one additional difference between factory systems and supply networks.—Or shall we rather say "category of differences", since this point encompasses several tiny and some significant differences. Obviously, Lean's ideal in terms of factory material flow, one-piece flow, is hardly transferrable to interorganizational material flow systems. JiT delivery is one measure to improve steady material flow between suppliers and customers, yet it sometimes seems to be subject to bogus implementation with suppliers producing to stock and merely delivering customers JiT with the quantities necessary. Here, many factors may play a role as to how well JiT (Lean) can actually be implemented in an interorganizational setting, such as geographical distance between supplier and customer and production lead time of the components supplied.

A significant difference lies in the purposeful location of the bottleneck. In a carefully planned factory system, the bottleneck does not emerge in a random place; and if a bottleneck shows up in an existing factory system there are effective measures to protect throughput and keep material flowing. Here again, things will be more complicated in a supply network.

Generally, it seems to be more difficult to manage and resolve bottlenecks in a supply network once the network has been established. Hence, a key lesson that should be taken from this paper is that bottleneck management should be considered right from the start of the planning of the supply network, i.e., that emphasis should be placed on *prevention*. Diligent planning of the supply network is more than choosing suppliers based on lowest price bids. Even more comprehensive approaches to supplier management (e.g., [24, 25]) often do not explicitly take bottleneck management into account. Firms that not only try to achieve smooth material flow within the boundaries of their factory system but aim to smoothen material flow across organizational boarders may thus enjoy competitive advantage.

References

1. Craighead CW, Blackhurst J, Rungtusanatham MJ, Handfield RB (2007) The severity of supply chain disruptions: design characteristics and mitigation capabilities. Decis Sci 38 (1):131–156
2. Li L, Chang Q, Ni J, Xiao G, Biller S (2007) Bottleneck detection of manufacturing systems using data driven method. In: 2007 IEEE international symposium on assembly and manufacturing, pp 76–81
3. Liu H (2011) A dynamic bottleneck-oriented manufacturing control system. Schriftenreihe: informationstechnische systeme und organisation von Produktion und Logistik, vol 13. Gito, Berlin

4. Hopp WJ, Spearman ML (2009) Factory physics. McGraw-Hill/Irwin, Boston
5. Haller M (2003) Cycle time management during production ramp-up. Rob Comput Integr Manuf 19(1–2):183–188
6. Goldratt EM, Cox J (2004) The goal. A process of ongoing improvement, 3rd rev. edn. North River Press, Great Barrington
7. Goldratt EM (2008) Standing on the shoulders of giants. Production concepts versus production applications. The Hitachi Tool Engineering Example
8. Moore R, Scheinkopf L (1998) Theory of constraints and lean manufacturing: friends or foes? http://www.tocca.com.au/uploaded/documents/lean%20and%20toc.pdf. Accessed 13 Jul 2013
9. Bicheno J, Holweg M (2009) The lean toolbox. The essential guide to lean transformation. PICSIE Books, Buckingham
10. von Bertalanffy L (1950) The theory of open systems in physics and biology. Science 111 (2872):23–29
11. Hall AD, Fagen RE (1956) Definition of system. General Syst 1:18–28
12. Harland CM, Knight LA (2001) Supply network strategy: role and competence requirements. Int J Oper Prod Manag 21(4):476–489
13. Harland CM, Lamming RC, Zheng J, Johnsen TE (2001) A Taxonomy of Supply Networks. J Supply Chain Manag 37(4):21–27
14. Gereffi G, Humphrey J, Sturgeon TJ (2005) The governance of global value chains. Rev Int Polit Econ 12(1):78–104
15. Provan KG, Kenis P (2007) Modes of network governance: structure, management, and effectiveness. J Public Adm Res Theor 18(2):229–252
16. Rowley TJ (1997) Moving beyond dyadic ties: a network theory of stakeholder influences. Acad Manag Rev 22(4):887–910
17. Barney JB (1991) Firm resources and sustained competitive advantage. J Manag 17(1):99–120
18. Beer JE (2011) Challenges and improvements in supplier management in the automobile industry. Master's Thesis, Rose-Hulman Institute of Technology
19. Forrester JW (1958) Industrial dynamics: a major breakthrough for decision makers. Harvard Bus Rev 36(4):37–66
20. Göpfert I, Braun D (2010) I want to hold your hand. Automotive Agenda (07/2010):85–87
21. Goldratt EM, Fox RE (1986) The race. North River Press, Croton-on-Hudson
22. Roser C, Nakano M, Tanaka M (2002) Shifting bottleneck detection. In: Yücesan E, Chen C, Snowdon JL, Charnes JM (eds) Proceedings of the 2002 winter simulation conference
23. Roser C, Nakano M, Tanaka M (2003) comparison of bottleneck detection methods for agv systems. In: Chick S, Sánchez PJ, Ferrin D, Morrice DJ (eds) Proceedings of the 2003 winter simulation conference
24. Falzmann J (2007) Mehrdimensionale lieferantenbewertung. Doctoral Thesis, Justus-Liebig-Universität Giessen
25. Janker CG (2008) Multivariate lieferantenbewertung. empirisch gestützte konzeption eines anforderungsgerechten bewertungssystems. Technische Universität Dresden, Doctoral Thesis. Gabler, Wiesbaden

On the Capitalization and Management of Infrastructure Assets: A Case from the North Sea on Its Natural Gas Export Pipelines

Eric Risa and Jayantha P. Liyanage

Abstract With the growth of demand for energy, natural gas was predicted to be a major energy commodity in immediate future. In the current setting, it would also act as a transitional energy source (due to its low emission) towards greener energy goals within EU countries. However, production and transportation of gas to various markets have also been met with different challenges, for instance due to economical as well as technical reasons. Norway is one of the leading natural gas exporters in the world, serving a large portion of Europe's energy demand through its subsea gas pipeline infrastructure. In light of not only global uncertainty, but also with respect to the potential market position, a major question is that what types of challenges are present in a specific critical downstream asset such as a subsea natural gas pipeline when striving to uphold the position as leading gas supplier to Europe. This is also a question of future opportunities in the ongoing energy debate despite the current economic conditions. This paper, based on a case study on a section of the large Norwegian gas transportation infrastructure, examines a wide range of multiple challenges, all of which could possibly challenge and pose a risk for the reliability and efficiency of gas supply from the Norwegian shelf. It elaborates on the current and future threats/challenges on the capitalization and management of the gas export pipelines from a longer-term perspective when striving for optimal asset availability in regards to future energy demand. The paper also elaborates on a specific scenario that would help the asset owners and other stakeholders to draw up a suitable strategy for long term value creation using a Strategy map.

Keywords Natural gas · Sustainable energy · Infrastructure · Business risk · Strategy map · Asset management

E. Risa
University of Stavanger, Stavanger, Norway
e-mail: ericrisa.mail@gmail.com

J.P. Liyanage (✉)
Centre for Industrial Asset Management, University of Stavanger, Stavanger, Norway
e-mail: j.p.liyanage@uis.no

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_43

1 Introduction

Utilization and management of infrastructure assets has brought major challenges in the modern industrial environment. Apart from well-known economical challenges, there appears to be various other factors that have explicit or implicit influence on the technical and operational decisions around infrastructures. These can vary from those that are more political in nature to the ones that have social, as well as regulatory relevance.

With the continuous growth of demand in energy markets, those infrastructures that are dedicated to distribution processes have received much attention lately. While this to a large extent is attributable to supply reliability, it appears today that there would be other regulating mechanisms that may shape up the utilization and management patterns in near future. This situation is particularly notable in oil and gas distribution infrastructure assets in Europe due to sensitivity and dynamics in downstream markets.

Even though the North Sea region is adequately equipped today with a number of natural gas distribution pipelines, the production potential coupled with European energy policy and other factors seem to be gradually gaining momentum to influence utilization and management processes in the future. Hence, there is an emerging need to review the current practices and strategies as well as to bring the regulating factors into spotlight that will shape up the future of gas infrastructure assets. Based on a case from North sea, this paper assesses the current situation from a broader perspective and elaborates on a suitable strategic path for an effective utilization and management practice for current infrastructures.

The written paper is based on an industrial case in North sea, where relevant data and opinions spanning across multiple disciplines and originating from leading specialists, has laid the foundation for the overall breadth in which asset management as a discipline seeks to acquire. Results were then derived on the basis of these finds, finally highlighting the critical factors relevant for the current topic.

2 Downstream Dynamics

2.1 Developments in Gas Markets

Global energy markets are changing, where uncertainty seems to be the common denominator. The dynamics of the downstream has been affecting the production and supply process gradually. Some of the principal observations in this regards includes:

- Supply strategy: Long distance transportation of energy has become increasingly common, resulting in a more globalized market with increasing market volatility in which trade flows of energy commodities are being directed to where the asking price is the highest. Consequently surplus energy is often exported instead of stored [1].

- Globalization: An effect of such volatility is that an event in one localized market will often affect the whole global energy market. Examples are abundant. Japan's nuclear Fukushima incident raised questions not only in Japan, but also in Germany, regarding the safety of nuclear power. Public opinion and politics lead to a sharp decline of nuclear output in Japan (−44.3 % in 2011) and Germany (−23.2 % in 2011) [2], thus forcing these nations to import energy from elsewhere in the global market. Japan's demand for non-nuclear energy resulted in LNG imports rising 11.2 % from 2011 to 2012 [4]. Subsequently in turn increasing regional LNG gas prices and diverting LNG trade flows to the region at the expense of areas where asking price and demand are somewhat lower, like Europe.
- Origin of European gas imports. Europe is currently the largest net importer of gas in the world, accounting globally for 46 % of total gas imports [5], where Russia and Norway via an immense natural gas pipeline grid are the top suppliers to the EU [6]. Approximately 48 % of European consumption of gas in 2011, was imported [7]. Of the imported gas, 2011 values saw approximately 68 % imported through pipelines and 32 % via LNG shipments [2]. Thus pipeline imports stand for a large portion in supplying Europe's gas demand and will do so in the years to come, especially as indigenous European conventional gas production declines. Natural Gas as an energy source has the past decade coincided well with European political ambitions of utilizing cleaner fuels. Along with other drivers, natural gas consumption within Europe steadily increased up till 2009 [6]. Between 2006 and 2009, European politicians began to question the reliability and trustworthiness of Russian gas, one of their main natural gas suppliers. In order to mitigate any future similar scenarios, politicians began to accelerate diversification of energy imports through amongst others LNG import terminals, gas storage buffers and subsidiary schemes for renewable energies. As a result, Middle Eastern countries like Qatar have built huge LNG vessel fleets to meet such demands.
- Financial Crisis: The financial crisis hit in 2008, growth stagnated, and directly impacting energy demands. Total energy consumption in Europe fell to pre-millennia levels, and so did also in turn the consumption of natural gas, down −9.9 % in 2011 from previous year [7]. At the moment, stagnation is still present within the region. 2012 saw gross national product in the European Union decline 0.6 %, where the forthcoming year is expected to yield a 0.4 decline (recently adjusted further down from 0.3) [8].
- US Shale gas: As Europe becomes more dependent on foreign suppliers of gas, the United States has seen huge energy independence benefits through its shale gas revolution. By utilizing new technologies, vast amounts of energy is suddenly available in the region, which is sold at relatively low prices. A direct result is a decreased local demand for indigenous coal, meaning that US surplus coal production needs buyers elsewhere. Germany for example, whom at the time is struggling with financial problems in the euro zone, see their energy needs being fulfilled by now available, cheap US coal (German coal import up 4.9 % from 2011 [9]). Even though carbon emission from coal is substantially

higher than natural gas, it has in some cases become more profitable to burn coal and buy carbon permits (sold under the EU Emissions Trading Scheme) for the surplus emission, than shift to other energy sources [10]. According to Bloomberg New Energy and Finance, a global energy research firm, German power utilities were set on average to lose €11.70 when they burned gas to make a megawatt of electricity, but to earn €14.22 per MW when they burned coal [11].

Looking back a few years, previous European prospects of natural gas was high, predicting a golden age for gas. These predictions are becoming increasingly uncertain, where coal may pose a larger role as transition energy for Europe towards renewable resources than what was previously anticipated.

2.2 Problem Domain from an Energy Provider's Perspective

Norway, the second largest supplier of natural gas to Europe supplies almost a fourth of European Natural Gas Consumption (107.6 bcm of 466 bcm estimated EU consumed in 2012 [5]). The gas is delivered through one of the world's largest subsea pipeline infrastructures, primarily supplying Germany, France and the United Kingdom. This places its main consumers in the middle of European environmental policies, in which are currently diverging mainly between gas, coal, nuclear, renewables and possible future indigenous shale gas [12].

In spite of deviations, overall European long term policy objectives are grounded in sustainability, security and affordability in order to ensure that the EU reaches its long term target of a 80 % reduction in green house gas emission by 2050. In addition, short term targets towards 2020 involve 20 % cuts in carbon emissions and energy use, along with increasing the share of renewable energy consumption from 8.5 to 20 % [12]. As targets are clear, the means in getting there are becoming increasingly complicated as the EU experiences financial problems.

What is certain, is that an energy transition within Europe is on its way, where renewable energy will stand as the energy choice of the future. The question is, what role is natural gas anticipated to play in EU's decarbonisation agenda? The outcome in this transition period lies not only within the European politicians in how they subsidize or add taxes and extra costs to various energy commodities, but also how suppliers of natural gas, such as Norway, positions themselves towards the core market customers and exceeding their expectations of a reliable, stable and competitive supplier of clean energy (compared to for instance coal).

3 Emerging Challenges from a Wider Perspective

In contrast to land based pipelines infrastructures, subsea pipelines of which rest predominately on the seafloor are subject to ever complex issues, these specific condition along with several other factors appear to centrally govern the future of

infrastructure asset management process. The level of observations with respect to the section of the gas export pipe selected in Norway, are as follows;

Remoteness: One of the main challenges is the shear remoteness of which a subsea infrastructure holds (depth, distance, temperatures etc.). Consequently, in comparison to onshore sections in which are often inspected every fortnight (or even more frequently), current technologies and their innate costs, often limit subsea routine inspections of pipelines to be held on annual basis', possibly also limited to a few selected legs. As such, understanding the implicit risks involved on the sea bottom, in order to not only mitigate, but also evaluate which lengths of pipeline should be prioritized for annual inspections, becomes a crucial factor when striving to limit the possibility of flow impeding threats occurring. Such threats could eventuate in pipeline downtime, resulting in not only revenue loss, but also impacting Norway's credibility as a stable and reliable supplier of natural gas.

Subsea pipeline threats have through the years been substantially monitored in order to fully understand the possible risks. Although most of the registered incidents are kept private within the industry, the PARLOC report presented in 2001 by The Institute of Petroleum, UKOOA and HSE, disclosed pipeline incidents in the North Sea which resulted in failures, with and without leakages, to oil and gas pipelines dating back to 1971 (All Norwegian main export pipelines are situated in the north sea). In the report, the dominating incidents where found to be: third party impacts 42 % (trawling and anchors), corrosion 27 % (internal and external), other 11 % (fittings, valves and unknown causes) and material 10 % (weld/steel defects).

Third party damage: Focusing on the largest issues within the Parloc report, namely third party impacts, approximately 70 % of all incidents in open water offshore pipelines (directly applicable to long export pipelines) were due to impact or anchors, and of those 70 %, 21 % led to a containment breach (all values excluding locations close to risers, platforms etc.). Consequently, random impacts, pose one of the greatest risks to the physical integrity of Norway's main export pipelines in open waters.

Trawling activities in the North Sea are abundant. In general, trawling equipment can either impact the pipeline or get caught on it. General impacts and pull over scenarios to the pipeline caused by trawling equipment, poses the largest long term wear risk over time for the external pipeline condition. As trawl board has increased the past years, it in no way poses any direct immediate threat to large export pipelines due to the low energy in which a travelling fishing vessel and its load carries.

One of the true challenges and threats lies within anchor impacts. Anchors, unlike trawling equipment, often belong to larger vessels such as shipping vessels which possess much larger mass and speed than trawling ships. Anchors can interact with a subsea pipeline in numerous ways, such as vertical impacts or more seriously in the event of a moving vessel's anchor snagging the pipeline. When snagging (getting caught) the kinetic energy from the moving vessel above will be transferred to the pipeline causing damage or even displacement until either the anchor chain breaks or the anchor is pulled over the pipeline. During any case, the

pipeline is most likely displaced, and in severe cases, resulting in local buckling which could or already have, developed into a possible rupture.

A large vessel of 100,000 tonnes will have an anchor weighing up towards 30 tones. Such an anchor can largely affect a pipeline, even if trenched 1–2 m below seabed [13].

As the economic activity in the North Sea increases, such as rising shipping traffic, the likelihood of such impacts also increase. Only recently, three large pipelines have been majorly affected by anchors (Kvitebjørn in 2007, CATS 2007 and Transmed 2008). In these instances, relatively big vessels with large anchors have hooked on to the pipeline and either majorly displaced or completely guillotined the pipeline(s) resulting in major downtime. On the basis of publicly available information regarding downtime caused by anchor incidents, along with predicted contingency response times, a 90 % confidence of (2 months < 51/2 months < 8 months) can be predicted for the time it takes until the pipeline is operational after a substantial impact. The natural gas value deferred or rerouted through one of the main export Norwegian pipelines within this timeframe can roughly be ball parked, on the basis of 2012 gas prices, to be between 1 and 4 billion Euro.

The true challenge of downtime mitigation lies in discovering such random impacts after they occur, but before leaks develop. More often than not, when large anchors have hit, local material integrity is compromised, but overall functional is not instantaneously affected. Over time, fatigue in the compromised area takes place due to gas cycles, eventually causing the pipeline to rupture.

In an environment where high inspection costs and low intervals are present, integration for more cost efficient equipment that can not only monitor whole lengths of the pipeline annually, but also inspect and monitor the pipeline at a higher frequency, becomes crucial.

As more suppliers to the gas market (LNG, shale etc.) and alternative energy sources are increasingly present, natural prices are bound to decrease, further accentuating the need for cost efficient operations in order to maintain the same marginal profit levels.

Competition: Russia, Norway's main competitor with a huge amount of available capacity, is further expanding its infrastructure further west, and its affect to Norwegian export can be questioned. Russia's main future strategies involve bypassing Eastern-European transit countries, thus feeding Western Europe directly, where their overall goal is to restore confidence to skeptic European politicians, reassuring them that Russian is a reliable source for natural gas energy. Russia is therefore most likely anticipating higher gas deliverance to Western European countries in the time to come.

Seeing as Western Europe is Norway's main customers, it may be likely that as the new Russian pipeline Nord Stream ramps up production, Norwegian gas supply may be affected. Preliminary studies and current export values seem to prove otherwise. Directly before the new pipeline was installed, analysis' predicted that as Russian gas exports through Nord stream, Norwegian export stays the same, whilst LNG imports fall. The report in addition concludes that the Russian pipelines Yamal (through Belarus and Poland) and Transgas (Ukraine, Slovakia) experience a

Table 1 European natural gas foreign supplies [1, 2, 3]

Import origin	2011	2012	y-o-y change
Norwegian total exports to Europe	92.8	106.6	+14.8 %
Russian exports to Europe	140.6	130.0	−8.1 %
European LNG imports have fallen 33 % year-on-year in the first half of 2012			

cannibalization effect of the newly introduced Nord stream. In fact, as natural gas consumption in Europe has fallen from 2011 to 2012, Norway has actually exported more on behalf of LNG and Russia.

Much of this growth can be due to a recent Norwegian effort to strengthen themselves in the natural gas market (Table 1).

Gas contracts: The European Union has the past decade had a vision in creating a common, open, competitive natural gas market, aligned with that of the UK's model. Such liberalized spot markets are gaining traction in a region where monopolized, often state controlled suppliers were firmly embedded, supplying natural gas on strict long-term, oil-indexed contracts. Such contracts were tailored in such a way that the buyer had to either take the supplied gas or pay a certain amount if the contracted gas was not needed. As Europe is turning towards market driven spot prices, it becomes evident that some gas suppliers still rely heavily on long term oil-indexed prices. Russia for example, had a mere 4 % of market priced spot gas prices in 2008, whilst Norway had 30 %. Statoil, Norway's main natural gas exporter, recently signed a 10-year gas supply deal to Germany based on spot gas prices, delivering 45 bcm more of its gas to the market. According to Statoil executive vice-president Eldar Saetre, their company is at the moment supplying more than 40 % of its European gas on spot terms [14]. As Russian energy policies have previously been heavily based on oil-indexed long term contracts, they seem persistent in sticking to this way of selling its gas, where Alexander Medvedev, Gazprom's export chief, stating in late 2012, that they will defend their system of long-term oil indexed contracts of all their energy [14]. It seems that the only way they would pose a large threat to Norwegian gas export, is by increasingly adapting to the Western European way in dealing with energy commodities, mostly in the form of spot prices, whilst moving towards a more transparent and reliable method of supplying stable gas flows to the market.

Even though the Norwegian export capacity is almost reached (107 of 120 bcm), one of the main overall challenges is that Norwegian gas production is projected to peak between 2015 and 2020 and decline thereafter if no further investments regarding searching for and developing new gas fields are made. Consequently, as future demand for Norwegian gas is to either stay relatively level, or surge, investments and efficiency aspects becomes increasingly relevant. The internal stakeholders of the infrastructural system hold a great deal of weight in ensuring the success of such goals.

Internal Stakeholders: In essence, all users (shippers) of the pipeline infrastructure system have to pay a tariff in order to send their gas to their designated

buyer. The tariff is paid to the owners of the infrastructure, namely the joint venture Gassled. To stimulate further development and offshore related activity, the Norwegian government regulates the tariff in order to ensure main gas profits are attained at the offshore fields and not in the transportation infrastructure (it should be noted that the tariff also takes height for any additional costs up to a certain level, such as increased maintenance costs). In 2009, 97.28 % of the ownership share in Gassled were either users of the infrastructure (Exxon, Statoil, Total etc.) or had direct ties to upstream interests (Petoro). During the past years, most of the major gas shippers in Gassled have sold their entire share (ConocoPhillips, Total, Exxon etc.) or heavily reduced it (Statoil). Buyers of the shares are long term investment companies such as Canadian pension funds. The ownership change was approved by the government, and the thought was, that in order to gain a more balanced risk of capital invested in offshore field and infrastructures, new financial owners should be included within Gassled. Such new owners would prefer other forms of financial returns than what the high risk, high pay oil and gas companies' pursue [15], namely stable returns over a long time period. As the Norwegian government is the somewhat sole regulator of their revenue through the tariffs they set, confidence (on the basis of previous Norwegian predictability) was given to the government, ensuring predictability for their investments.

One of the main changes is that the majority of Gassled owners are now not direct users of the system. They have hardly any direct interest to the upstream market. Implications of such changes to strategies are for one that the new owners have no need to keep costs down. On the contrary, according to Alexander Engh, consultant in Deloitte, the new owners will have an incentive to increase for example maintenance costs, thus ensuring higher regularity/availability giving higher profits [16]. These extra costs will be deferred to the users of the system, which are not the owners anymore since they are not gas shippers. This could become a positive aspect in relation to the wellbeing of the infrastructure pipelines, ensuring a high quality standard, but a negative aspect regarding economic efficiency and above all, competitive natural gas prices. Such overall excess spending would not only lead to dissatisfied shippers (users), but as a secondary result, social economic loss may occur through cost ineffectiveness. The past decade since the regulation system was implemented, there has been a relative balance between user- and owner-interests within the infrastructure. This balance has changed, and implications are present within the organization.

Subsequently following the last major oil and gas operator selling its' Gassled holdings, the government (a year later) releases a consultation paper in January 2013, suggesting to change the tariff levels in the Gassled infrastructure network. Their reasoning in lowering the tariffs was based on the take, that a reduction in transportation costs would stimulate and provide further incentives for oil and gas companies to further invest in exploring and developing new fields, especially further north in order to sustain Norwegian competitiveness currently and in the future. Should the tariff reductions be implemented, transportation costs for users will be heavily reduced, directly impacting Gassled's revenue stream negatively. According to Norwegian newspaper Aftenbladet, the newly investment owners of

Gassled feel tricked by the Norwegian government in what they were thought to believe was a safe investment giving a minimum expected return of 7 % on capital invested. They now anticipate their return of halving, transferring values of approximately 5 billion Euro from Gassled to the gas shipper, namely the oil and gas operators. This may then imply that the shippers are the winners of such a cut, due to the lower transfer costs, increasing their profitability of not only current, but also future projects. Moreover, the government would also gain profit increases, through tax incomes of future found gas resources, more jobs, and in general social growth [15]. Furthermore, should the new owners have their "stable returns" on investments reduced, their willingness to ensure other relevant stakeholder's wellbeing may be affected negatively.

A conflict of interested thus arises, where the investment owners are seeking a 7 % or more return on investments on midstream level, whilst the state and oil and gas operators want to specifically increase production and sales in upstream location. Such elementary drivers for profit will negate overall efficiency and introduce conflicts of interests between financial stakeholders within Gassled, oil and gas operators within Gassled, the government and infrastructural operator Gassco. Even many of the oil and gas companies see negative effects of such an imposed tariff. By decreasing Norwegian predictability and stability, future investors might opt for other investments in other countries, possibly resulting in the oil and gas operators yet again are left with ownership in Gassled of what they deem, low interest investments [15] (Oil and Gas companies build the pipelines, and if/when the new pipelines are included in Gassled, the companies receive an ownership stake corresponding to the pipeline's value).

In addition to the challenges previously discusses, other global threats such as increasing: ship traffic, cyber attacks targeting energy infrastructures, shale gas development and LNG, does not make the situation any easier. All of which could possibly challenge and pose a risk for the reliability and future demand for Norwegian gas.

4 Resolving the Future: A Strategy for Long-Term Value Capture

As main threats such as anchor impacts, competitors, internal stakeholder issues and impending tariff changes are present, pathing the way through these challenges whilst ensuring long term value through overall strategies can be a daunting task. Nevertheless, an attempt in this context is illustrated in Fig. 1. The following sections sees to further describe how and why this strategy map was derived and further describe the content.

The past decade has seen a huge shift in how organizations create value. Kaplan and Norton opened the eyes of businesses. They introduced the fact that more than 75 % of a firm's market value is derived from intangible assets, and therefore new

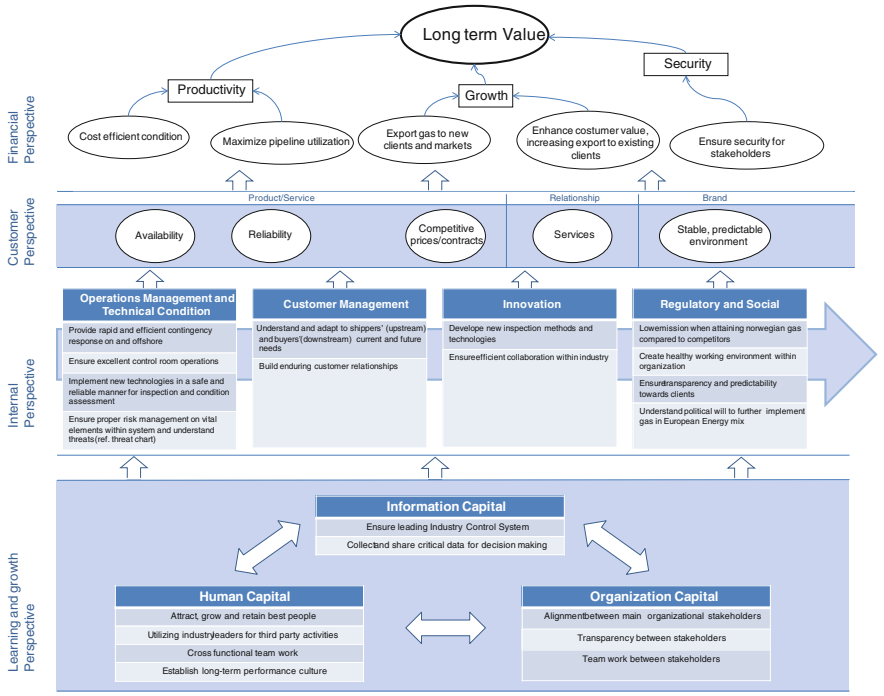


Fig. 1 A strategy mapping effort to elaborate a long-term value capture solution under the present dynamic conditions

ways of capturing these values were possible through their balanced scorecards. As organizations implemented Kaplan and Norton's balanced scorecards, they saw a further need for a more powerful way of implementing their strategies in an optimal fashion. What most organizations experienced, was that they failed to implement their new strategies within the organization. Employees would read the strategy statements and visions, but would fail and implementing them within their work place. In fact, 70–90 % of organizations failed to realize success through their newfound strategies [17].

In 2004 Kaplan and Norton introduce their new tool: Strategy maps, which has been receiving rave critics, being described as innovative and important as the balanced scorecards. "The strategy map provides the missing link between strategy formulation and strategy execution" [17]. One of the main criticisms Kaplan and Norton have received in light of their balanced scorecards (in which strategy maps is based on), is the fact that they have a too narrow stakeholder focus [18], thus devaluating other relevant stakeholders within the strategy. This issue has become especially criticized by stakeholder theorists who believe that there are other important parties other than the immediate needs of shareholders and stockholders in which focus should be directed towards in gaining long term value.

Kaplan and Norton's Strategy Map is sectioned in four segmented focus areas where the first and last and their applicability to the asset are specifically described below.

- Financial Perspective: In order to utilize Kaplan and Norton's strategy map whilst in addition valuing the criticism by stakeholder theorists, a third focus objective is incorporated into the map (in addition to productive and growth) called "security". Elements such as financial security, environmental security and social security becomes objectives in which facilitate long term value creation. For example: Ensuring long term environmental security for local inhabitants will yield positive public opinion, resulting in consent form the masses. Public opinion should not be underestimated, and the power of the masses has on multiple occasions throughout history changed the courses of large financial investments. Or it could be the ensuring of social security in the form of taxes in which goes to the government and hence the public. By adding stakeholder "security" as an objective, the strategy incorporates long term stakeholders in which then in turn ensures long term sustainability within not only the financial setting, but also settings where other relevant stakeholders act within.
- Learning and growth perspective: As tangible challenges regarding global markets, third party impacts, customers, competitors etc. In no way should be treated tertiary, the underlying intangible elements: Information Capital, Human Capital and Organization capital (encompassed by Gassled, Government etc.) are the glue holding the organizational structure together. Common theory dictates that the three intangible asset groups are interconnected and must together complement and be mutually supportive to one another in order to attain success. Changing a factor in one of the three will affect the two others, thereby requiring mutual strategic attention towards all of the three segmentations. Focus towards all three requires diligent awareness in regards to trade-offs and compromises being made between them, in order to attain the overall strategy [19].

Regarding the strategy, specifically for the new financial investment owners in Gassled, their incentives may be solely towards gaining as much profit as possible for its confined shareholders through the tariff. According to Kaplan and Norton, long term values are targeted to be just this, namely value to its shareholders. But by in addition, introducing a "security" element within its top value gaining elements, they are for example indirectly forced to think of what investments they make could gain social security within Norway, thus aligning the strategy to other beliefs and values extending past the internal local stakeholders within an organization in order to fully stimulation long term growth. By focusing on such matters, they involve themselves to a higher degree within upstream planning which could lead to an increase in wealth for not only them, but the nation in general by providing job security and further employment, basically aligning themselves with the government and somewhat upstream owners within Gassled. Neglecting the security element within the strategy, investment companies may possibly overinvest in the infrastructure, shifting its increased cost to the users (oil and gas companies)

through an increase in tariff. This would in turn reduce the profit margin for these companies, thus stagnating future investments, negatively affecting social benefits, like jobs, profits through future find petroleum taxes etc.

5 Conclusion

As the rest of the world clings tightly to, and expects increased consumption of natural gas due the benefits of the commodity, Europe will most likely follow suit in due time as their economy gets back on track. But even relatively independent of which way European demand for natural gas sways, Europe cannot, and will not most likely easily wean itself off gas as an energy source. Even if the energy commodity was to lose ground in the power utility sector, there will inevitably be a need for natural gas due to an already present and robust infrastructure supplying natural gas to homes and industries across the continent. In this sense, the potential for value creation as a natural gas supplier will be evident in not only future short term perspectives, but also long term.

The main common issue is that as the world has become ever more globalized, natural gas suppliers have to become increasingly competitive in order to sustain or gain market shares. This becomes especially evident in a financially troubled Europe where energy consumption is decreasing.

Norway, with a lean, effective infrastructure, is presently aggressively adapting to market demands, and may soon overtake Russia as the main supplier of natural gas.

But in order to further increase its export capacities, uphold its availably and ensure predictability, multiple inherent and external challenges must be managed effectively as to sustain Norway's reputation as a prominent contender in supplying the present and future gas market. By creating a strategy road map in order to effectively deal with present and arising challenges, the infrastructure and its stakeholders may together ensure that further growth and prosperity can be created not only amongst its internal stakeholders, but also the people of Norway of which whom the natural gas resource truly belongs to.

References

1. DG Energy (2012) Quarterly report on European gas markets, vol 5, issue 4
2. BP (2012) BP Statistical review of world energy, June 2012
3. BP (2013) BP Statistical review of world energy, June 2013
4. LNG world news (2013) [online] Available at: http://www.lngworldnews.com/japan-december-lng-imports-climb-7-4-percent/. Accessed Jan 30 2012
5. Enerdata (2013) 2,2 % drop in European gas consumption in 2012. [online] Available at: http://www.enerdata.net/enerdatauk/press-and-publication/energy-news-001/22-drop-european-gas-consumption-2012_17554.html. Accessed June 4 2013

6. Eurostat (2012) Gross inland consumption in EU-27. [online] Available at: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php?title=File:Gross_inland_consumption_in_EU-27_2011_in_million_tonnes_of_oil_equivalent_(Gross_Calorific_Value).png&filetimestamp=20120529132738. Aaccessed Feb 4 2013
7. Gazprom (2012) Gazprom to continue reinforcing its standing in Europe. [online] Available at: http://www.gazprom.com/press/news/2012/april/article134112/. Accessed May 18 2013
8. Landre E, (2012) EU nedjusterer vekstutsiktene. [online] Available at: http://e24.no/makro-og-politikk/eu-nedjusterer-vekstutsiktene/20365096. Accessed May 3 2013
9. Bloomberg (2012) Merkel's green shift forces germany to burn more coal. [online] Available at: http://www.bloomberg.com/news/2012-08-19/merkel-s-green-shift-forces-germany-to-burn-more-coal-energy.html. Accessed Jan 30 2013
10. Burgess J (2012) Coal consumption increases in the EU: is the carbon trading scheme a failure?. [online] Available at: http://oilprice.com/Energy/Coal/Coal-Consumption-Increases-in-the-EU-Is-the-Carbon-Trading-Scheme-a-Failure.html. Accessed Apr 15 2013
11. Andresen T, (2012) EON loses as RWE's coal plants win Germany's green shift: energy. [online] Available at: http://www.bloomberg.com/news/2012-12-07/eon-loses-as-rwe-s-coal-plants-win-germany-s-green-shift-energy.htm. Accessed Jan 30 2013
12. Bjørnson R, (2013) Natural gas- key to transform Europe's energy system – presentation during 1st EU Norway energy conference. Available through: European commission website: http://ec.europa.eu/energy/gas_electricity/events/20130305_norway_energy_conference_en.htm. Accessed Apr 13 2013
13. HSE (2009) Guidelines for pipeline operators on pipeline anchor hazards. [online] Available at: http://www.hse.gov.uk/pipelines/pipeline-anchor-hazards.pdf. Accessed Apr 12 2013
14. Adomaitis N (2012a) Norway challenges Russia with new European gas pricing. [online] Available at: http://www.theglobeandmail.com/report-on-business/international-business/european-business/norway-challenges-russia-with-new-european-gas-pricing/article5472474/. Accessed Apr 9 2013
15. Rovik TM, Fossøy K (2013) Fra stabilitet til statlig opportunisme. Stavanger Aftenblad May 24 2013
16. Lewis HØ (2011) Må legge til rette for nye Gassled-eiere. [online] Available at: http://www.aftenbladet.no/energi/Ma-legge-til-rette-for-nye-Gassled-eiere-2894849.html#.UV1MQxziks4. Accessed Apr 4 2013
17. Kaplan RS, Norton DP (2004) Strategy maps: converting intangible assets into tangible outcomes. Harvard Business Press, Boston
18. Flak LS, Dertz W (n.d) Stakeholder theory and balanced scorecard to improve IS strategy development in public sector. [online] Available at: http://web.bsru.ac.th/~thanarat/IRIS2028-1109.pdf. Accessed May 5 2013
19. Perrow C (1999) Normal accidents: living with high-risk technologies, Chapter -3: Complexity, Coupling, and Catastrophe

Current Status and Innovative Trends of Asset Integrity Management (AIM): Products & Services in the Norwegian Oil and Gas Industry

Oluwaseun O. Kadiri, Jawad Raza and Jayantha P. Liyanage

Abstract Owing to commercial opportunities in the Norwegian Oil & Gas (O&G) sector, there is a growing demand for new products and services for Asset Integrity Management (AIM). In general, the market demands greater degree of innovation in the field of AIM seeking means to simplify complex work processes and at the same time to have a better understanding and awareness of inherent risks. The trends for innovative AIM products and services can always be challenged due to such factors as cost, organizational capacity, technological capacity as well as underlying business growth potential of the innovation. Other barriers may arise from financial constraints, regulatory requirements, non-proven technology and clients' conservative perspective to invest in new and revolutionary products. Keeping such challenges in mind, this paper attempts to map the current status and future trends of Asset Integrity Management (AIM) products and services in the Norwegian O&G industry. This paper highlights the status and gaps through a thorough literature and market survey to identify the type of AIM services and products within AIM for Norwegian O&G assets. It also highlights new emerging trends from AIM contractors/service providers to align their products to match with the new asset integrated operational environment, such as Integrated Operations (IO).Interestingly, there appears to be less innovation in the industry despite the fact that the age of the industry is increasing. Some of the reasons include limited knowledge and competencies, operators and regulatory bodies' conservative attitude towards new technologies. This attitude regulates the development and deployment of AIM due to its sensitivity in terms of managing asset related uncertainties and vulnerabilities.

O.O. Kadiri (✉) · J.P. Liyanage
Centre for Industrial Asset Management (CIAM), University of Stavanger, Stavanger, Norway
e-mail: oo.kadiri@stud.uis.no

J. Raza
Department of Maintenance and Operations, Apply Sørco, Sandnes, Norway

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_44

Keywords Asset integrity management (AIM) · Products & services · Innovation · Trends · Barriers · Operational integrity management (OIM) · Technical integrity management (TIM)

1 Introduction and Background

The safety and integrity of assets have always been a major concern for operators. This is because not only can a well-managed asset integrity program help operators identify and reduce safety risks before they escalate, asset integrity can also play a major role in achieving higher operational excellence and extending the remaining useful life of ageing assets [1]. History reveals that lack of proper management of the assets can have negative effects such as personal injuries, loss of production, catastrophic events that can lead to loss of life, reduced equipment reliability and environmental impact. Asset Integrity can be defined as the ability of an asset to perform its required function effectively whilst safeguarding life and the environment [1]. Many AIM service providers/contractors are assisting operators to ensure that their assets function safely, effectively and economically throughout its life cycle. In order to gain full benefits of an effective AIM in a dynamic operating environment, it is essential that all stakeholders have a consistent and a unified understanding of what the essentials of asset integrity are and how these can be applied in their day to day operations. This is often cited among the most significant challenges in achieving an integrity culture within an organization. The implementation of asset management practices within an organization enables it to see tangible benefits such as higher safety standards, lower operating costs, longer asset life, improved asset performance, greater reliability, enhanced environmental support and better informed investment strategies.

The concept of AIM products and services in the Norwegian industry has been continuously changing over the years. Until the late 1960s the integrity of the design and operational safety of offshore platforms was largely the responsibility of the owner-operators who used a variety of industry and in-house standards and methods mostly visual inspection. Accidents did not receive much publicity outside the industry because few were lost and at the time, there was little concern about pollution. The Ekofisk platform Bravo blowout in the North Sea that occurred in 1977 was one of the major accidents that have a profound effect on the way the offshore industry does business in Norway and worldwide [2]. This accident created a higher level of government involvement towards improving safety of the O&G related activities. Several requirements were imposed by the regulatory bodies to perform detailed platform and operational probability risk assessments in order to demonstrate the overall reliability and to meet minimum acceptable safety and reliability criteria for the facility. Over the past 30 years Norway has moved away from a strict prescriptive approach to a more performance-based approach for regulating offshore O&G facilities. Performance-based regulations guide operator

companies to determine the best way to achieve operational and technical safety targets. The regulatory requirements from Norwegian Petroleum Directorate (NPD) are general in nature and primarily specify the conditions that must be met to be in compliance with the requirements. Within this framework operators have the freedom to choose practical asset integrity solutions along with the responsibility to ensure higher levels of compliance. To avoid misunderstandings about the requirements for compliance the Det Norske Veritas (DNV) "Offshore Standards" publications define the technical requirements and acceptance criteria [3]. Currently most operators have outsourced the management of their physical assets to service operators so that they can focus on their core business which is production without having to compromise the integrity of their assets while trying to meet organizational objectives. This concept has also evolved over the years. The Norwegian Continental Shelf (NCS) is advancing more and more into deeper sea operations with significant subsea developments where they have to face more harsh conditions in remote locations (Lokko et al. 2012). This therefore calls for the need of more robust products to control and manage assets remotely. Initially AIM service provider companies deliver tools for maintenance which is known as the "product concept" but now the industry is moving toward the solution concept where these service providers offer solutions to assist operators in enhancing asset performance to become more productive and competitive. In this context the service providers not only focus on the product they are delivering but also the quality of the relationship that they have with their client since it is no longer a one time delivery. In this environment, the purpose of this paper is to identify how service providers are helping to maintain this relationship by providing asset management solutions, the challenges they face in developing innovative products and the drivers that enables them involve in innovative projects.

2 Methodology

This study was carried out through thorough analysis of literature and a market survey. The survey consisted of four case studies of AIM service providers operating on the NCS. The aim of the survey was to identify and highlight recent AIM products and services and the innovation trends on the NCS. The data was collected through questionnaire-based interviews with experts in the field of asset integrity. The scope of the survey included questions about product and services these companies offer, driving factors for developing new products/solutions, innovation processes as seen from AIM service provider's view, feasibility factors and innovation barriers in developing innovative AIM solutions for NCS Operators. The results and deductions from the survey may therefore be limited due to quality of data obtained from the case studies. During the scheduled interview sessions with the experienced management professionals, their views were recorded and the data was analyzed to highlight current status, trends and challenges in embarking on innovative AIM products and services.

3 AIM Status on the NCS

Integrated operations (IO) have been the new face of optimizing AIM in the Norwegian O&G industry. IO is a term used for the implementation of ICT in the O&G industry to combine work processes, technology and organization together in a seamless way with the aim of improving production operations and support. This concept was first introduced in the O&G industry by the Norwegian petroleum industry making them the pioneer of this concept for petroleum related activities [4]. The most striking part of IO has been the use of always-on videoconference rooms between offshore platforms and land-based offices. This includes broadband connections for sharing of data and video-surveillance of the platform. This has made it possible to move some personnel onshore and use the existing human resources more efficiently. Instead of having e.g. an expert in equipment condition monitoring on duty at every platform, the expert may be stationed on land and be available for consultation for several offshore platforms. It's also possible for a team at an office in a different time zone to be consulting the night-shift of the platform, so that no land-based workers need work at night [5]. Splitting the team between land and sea demands new work processes which together with ICT is the two main focus points for IO. Tools like videoconferencing and 3D-visualization also creates an opportunity for new, more cross-discipline cooperation. For instance, a shared 3D-visualization may be tailored to each member of the group, so that the geologist gets a visualization of the geological structures while the drilling engineer focuses on visualizing the well. Here, real-time measurements from the well are important but the down-hole bandwidth has previously been very restricted. Improvements in bandwidth, better measurement devices, better aggregation and visualization of this information and improved models that simulate the rock formations and wellbore currently all feed on each other. An important task where all these improvements play together is real-time production optimization. Optimizing Asset Integrity Management with Integrated operations involves adapting to new changes in work processes and technologies. In order for this process to be successful the organization needs to be flexible and willing to change their work methods to fit the new work processes that are being implemented.

3.1 Maintaining Technical Integrity

From literature studies and interviews from experts in this area, it has been observed that from the inception of the petroleum industry on the NCS in early days, the method of maintaining technical integrity was through the run-to-failure management system. Run-to-failure is a reactive management technique that waits for machine or equipment failure before any maintenance action is taken. The main reason behind this was lack of factual data that quantifies the actual need for repair or maintenance of the plant, equipment and systems. This is because the major

expenses associated with this type of maintenance management are: high spare parts inventory cost, high overtime labor cost, high machine downtime, and low production availability (Mobley 1990). It was later realized that maintenance scheduling can be carried out effectively based on statistical trends of the performance or failure of the plant/equipment. The most common method of maintaining technical integrity presently on the NCS is the through preventive measures. The transition to this method of maintaining technical integrity has been adopted by all operators on the NCS. The concept of Preventive Maintenance (PM) management method is that maintenance tasks are based on elapsed time or hours of operations i. e. time-driven (Mobley 1990). This is generally done using the statistical life of equipment which is also known as mean-time-to-failure (MTTF) or bathtub curve. A bathtub curve indicates that a new machine has a high probability of failure, due to initial installation problems, during the first few weeks of operations. Following this initial period, the probability of failure is relatively low for an extended period of time before it then increases again with time. Based on bathtub distribution a machine should be maintained or modified on a schedule based on MTTF statistic. All PM management programs assume that machines will degrade within a time frame based on its classification. One important parameter for identifying the suitable PM strategy is how quickly or slowly the equipment degrades and how detectable the failure mode/mechanism is. For example a single stage split case centrifugal pump will usually run 18 months before it must be stopped for maintenance that means using this method of maintenance management the pump must be maintained at 17 months of operation to prevent total breakdown before repair. This management method is better than the previous method used because it is less expensive. The downtime of equipment using this planned management is lesser and planned but it has its own disadvantages. The disadvantage of this method is that it only considers MTTF but the problem is that it might end up in unnecessary repair or catastrophic failure in between these times. In the example given earlier, the pump may not need to be maintained after 17 months. Therefore the labor and material used to make the repair are wasted. On the other hand, the pump could fail before 17 months forcing the management to use run-to-failure techniques. The preventive maintenance method has been developed and adopted by the NCS operators is using integrated operations to optimize technical integrity through real time data condition monitoring and remote diagnostics. In summary the development trend of maintaining technical integrity of offshore assets on the NCS has been gradually from corrective maintenance or “no maintenance” in the 1960s to PM to real time condition monitoring in 2000s. Figure 1 shows the gradual development trend of maintaining technical integrity on the NCS.

3.2 Maintaining Operational Integrity

Operational integrity on the NCS is a major source of production performance and also the source of most safety related issues. The requirement for operational

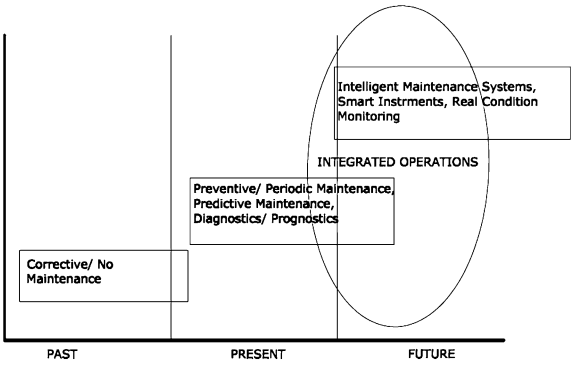


Fig. 1 Technical integrity management development trend on the NCS

integrity is usually ergonomics which includes the working environment and the clarity of the information available to operate. Operational integrity on the NCS has gone through different phases of development over the years. Initially on the NCS operational integrity was achieved through in house expertise and know-how. This method of ensuring operational integrity has its own consequences because the O&G industry is a high risk industry and high risk organizations do not have the luxury to learn by trial and error [6]. The consequences of error in these organizations are often so great that could result in loss of lives and equipment. Also the time shift of personnel is important so as to increase personnel alertness on the job. If this is not properly looked into error would be prevalent when the alertness of the personnel is low. Currently on the NCS the way of ensuring operational integrity is through the use of simulator training, in house experts, in-sourcing and outsourcing. During the training to ensure operational integrity, the simulators that are used models different scenarios on an offshore platform. The trainee involved use the simulators to learn how to carry out operations procedures and respond to critical scenarios without actually having any negative effect because the environment is entirely a virtual environment. This method of ensuring operational integrity is better than the past method but also has its limitations. The major limitation of this method is the lack of knowledge. Even though different scenarios are designed in the simulators for the trainee to learn during training, in real life operations there are still scenarios that would occur which are entirely going to be new to the operator. Making a wrong decision in scenario could result in a catastrophic event. This has been a reason while the operators are now finding better ways of ensuring operational integrity. The new method that is being developed and adopted by some operators is the use of the integrated operations platform to created real-time experts online support and remote operations. It can be said that the capacity development of the future in order to maintain operational integrity, would be the use of mobile workers with real time global experts' access. In summary the development trend of ensuring operational integrity during operations on the NCS has been from in-house expertise to the complex collaborative solutions through remote operations/online support. Figure 2 shows the capacity development trends for ensuring operational integrity on the NCS by capitalizing on new capabilities.

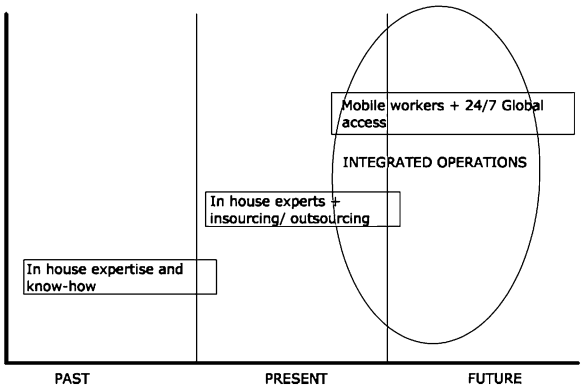


Fig. 2 Capacity development trends of operational integrity on the NCS

4 Results and Discussion

From the survey carried out, a common trend for driving mechanisms of innovation was observed for all four AIM service providers on the NCS. New and cost effective technological AIM solutions are seen to be followed by market demands and client specification and preferences (see Fig. 3). Market demand remains as most influential driver for innovation because Operators demand for continuous optimization of the asset in order to comply with the regulations. From the survey, company A and D have the same response to what drives innovation for them which is what is common for others except employee initiative. Company B on the other hand has client specification as one of their drivers of innovation. According to the representative of the company during the survey said that they push the boundary of AIM further based on the operators preferences and specification. This

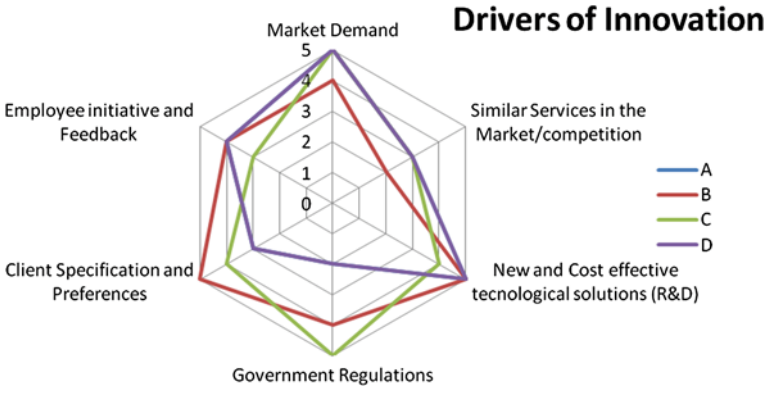


Fig. 3 Drivers of innovation for AIM product and services on the NCS

also because when the operators make specifications and preferences, they also are willing to participate in the project that would bring about innovative AIM products and services. Company C observed that over the years the main factor that has also drives them among other factors is government and authorizing body regulations. When these regulations are put in place then all operators would need to work within that boundary and still make the most opportunities they can within that boundary. This in turn make operators look for better AIM products and services that can meet their organizational objectives and is still within the specified regulations. Also, the new AIM solutions should contribute to reducing cost and increasing safety and efficiency which every business owner wants when running their business.

Figure 4 shows the barriers of innovation, as seen from the result of case studies, on the NCS within the area of AIM products and services. Innovation barriers that are evident on the NCS where identified and the way companies responds to these barriers where evaluated. It was noted that common barriers that all four service providers on the NCS suffer from is lack of skilled personnel and financial constraints. The companies studied have these factors in common but reacted to it in deferent manners this is shown in Fig. 4. From the survey, company A has financial constraints as a major barrier but from the interview with the personnel at the company, he emphasized that if there is demand and appropriate capital, the barrier of competencies could be tackled through collaboration and outright buying of the needed competencies. However, company B recognizes lack of skilled personnel as a major barrier towards innovation because innovative projects would need highly skilled personnel to implement which according to this company is scarce in the Norwegian oil industry irrespective of remuneration. This situation in company B goes for company C as well.

Innovation feasibility factor in this context can also be seen as success factors from AIM service providers' point of view when engaging in innovative projects. These factors are what companies consider to see if the innovative project is worthy in long

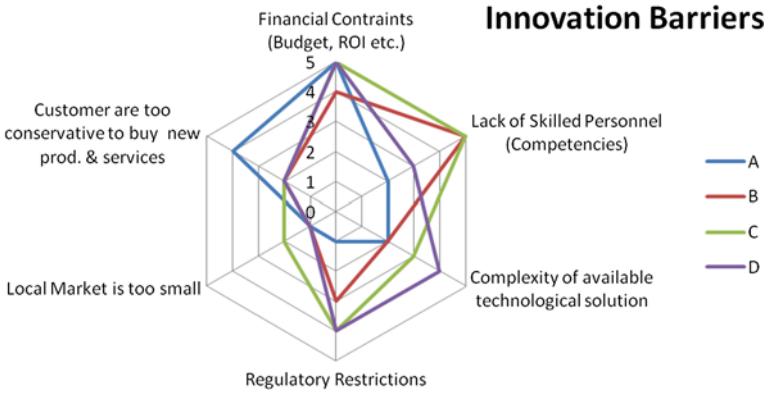


Fig. 4 Innovation barriers for AIM product and services on the NCS

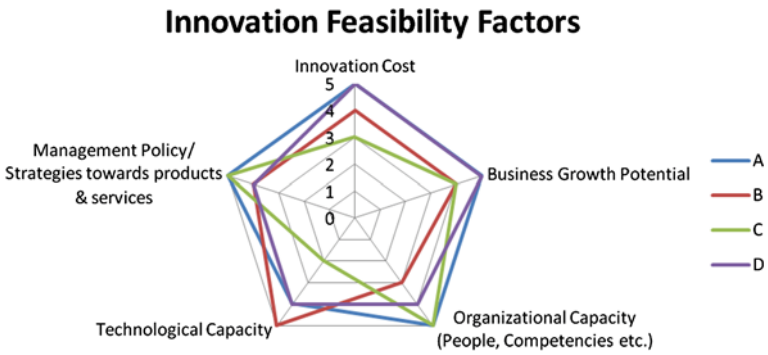


Fig. 5 Innovation feasibility factors for AIM product and services on the NCS

term. From the survey results, company A sees innovation cost, management policy/strategies towards products and services, business growth potential and organizational capacity as major factors while company B considers technological capacity as a most influencing factor above all. Company C reflects management policy/strategies towards products & services and organizational capacity as key factors whereas for Company D innovation cost and business potential are vital factors. It is noted that the one of the fundamental factors to be considered in this context is management policy and strategies towards new AIM product and services development. This is because there are different scenarios that pose the different challenges and the most important concluding factor is the way these Companies react towards these challenges. This could be embracing innovation, rejecting innovation or been indifferent about the status of innovation in the company (Fig. 5).

From the survey results, it was noted that irrespective of market demands, management policies and organizational strategies are significant factors that drive innovation. This is because despite the obvious opportunities, the choice still lies in the hand of the higher management. It was observed that most operator companies may be to some extent reluctant to be a part of innovative processes as this may result in organizational and changes to existing management and work processes. Anyhow, in case of any serious incident or a new government regulation leave them with no choice but to seek for new innovative products to ensure their asset's integrity. It was said by one of the senior staff at an Operator Company that operates on the NCS that "Most times the company is usually contented with their profit especially when production is stable and there is no threat from the regulatory authorities and the government". The amount of profit which is also the aim of the business has an effect on the way company drives cost efficiency. This attitude can also make operators to become stagnant in their demand for more innovative products. At the SPE conference held at Houston in February 2013, it was said by the Chief Technology Officer of a multinational O&G company that "The Oil and gas industry is one of the least innovative industries in the world in comparison to other industries such as the aviation industry, medical industry, communication industry etc." The leap in technological advancement in the O&G industry is quite little despite the fact that the

industries he mentioned too can be categorized as high risk industries. These O&G companies most of the time don't have rivals in terms of profits and that also tends to make them rest on their oars. Also having consistently abundant funds make operator companies to continue to pay exorbitant prices for product and services that they're familiar with rather than embrace new or innovative concepts that would be cost effective and efficient. On the side of the AIM service providers, there were also some trends that were also noticed. Although it is logical for customer demand to drive innovation, some service companies drive innovation themselves without customer demand. This was noticed by one of the major service providers on the NCS. The company representative during the interview categorically said that their company is "technologically driven". For this company, their belief is that these innovative solutions could be made ready and then they would make their client see the need to having these products or solutions. Also it was noticed that other less competitive companies do not see technological capacity as a main challenge. This normally is a point of concern for them but they consider collaboration and buying technology from a third-party as a very viable solution for them. This is because they don't see themselves as "technologically focused" but "customer demand focused". Also it was noticed that some service providers create revolutionary products from "adhoc processes" which is not usually a norm for most service companies. Most services companies have a structured process to incrementally (evolution) develop their products and services. Organizational capacity is also one of a major challenge towards developing innovative solutions in for the AIM service providers. The Norwegian O&G industry generally lacks human competencies than most O&G industries in the world which tends to influence the provision of specialized skill.

4.1 Identified Gaps and Barriers

From the study, the following gaps and challenges have been identified.

1. Limited Knowledge
2. Management strategies and organizational policies.

One of the major gaps which influence the development of innovative products and services for the Norwegian O&G industry as highlighted above is *limited knowledge*. Limited manpower available in Norway has become a challenge for both operators and AIM service providers. A new trend in bridging this gap is relying on competencies within Europe, which itself poses new challenges. The required competences involve both engineering and Information Technology (IT). The treasure in the future of innovative work processes would include adequate data together with human competency to interpret these data to produce insightful basis for decisions and solutions. *Management strategies and organizational policies* is a gap that needs to be filled. Most leaders have never learned how to be innovative and how to lead an organization so that it becomes more innovative. They may understand that they have a key role in innovation, but they do not know how to

systematically generate new and better solutions. They also do not know how to reinforce the right innovative skills for their direct reports and teams. Therefore people who have a proper grasp of innovation should be made to lead and make policies and decisions that would drive innovation in the organization.

5 Conclusion

AIM in the Norwegian oil and gas industry has seen to be continuously improving over the years. The implementation of the integrated operations (IO) scenario has made it possible to optimize asset integrity by improved production, extended asset life, improved safety and reduced cost. This platform has also opened new challenges for advanced asset control and optimization. As a result of this technological revolution, AIM service providers are facing challenges not only in upgrading their existing AIM products and services but also to developing new and innovative solutions. These challenges are largely affected by Operators' conventional behavior, regulations, financial constraints and lack of competent personnel. These factors, to some extent, seem to regulate the development and deployment of new and innovative AIM solutions. Therefore, more could still needs to be done to improve and overcome the innovation barriers which today are major obstacles the Norwegian O&G industry from moving into the next phase of IO development and implementation.

Acknowledgments The Authors wish to deeply thank all who participated in the survey and provided their useful input in completing the survey.

References

1. Rao RA, Rao SS, Sharma T, Krishna RK (2012) Asset integrity management in onshore and offshore-enhancing reliability at KGD6. In: SPE oil and gas India conference and exhibition. Mumbai, India
2. Visser RC (2011) Offshore accidents, regulations and industry standards. In: Proceedings of the SPE Western North American regional meeting held in anchorage, Alaska, USA, pp 7–11 May 2011
3. DNV (2010) Safety principles and arrangement. Det Norske Veritas, Offshore Standard DNV-OS-A101, October 2010
4. Erstad C (2011) Present and future technical integrity management practices for integrated operations. In: Master thesis faculty of science and technology, University of Stavanger
5. Hauge ST (2011) A study of integrated operations on the norwegian continental shelf. A master thesis at the University of Stavanger
6. Roberts K, Gargano G (1990) Managing a high-reliability organisation: a case for interdependence. In: Von Glinow M, Mohrman S (eds) Managing complexity in high technology organisations. Oxford UP, New York, pp 146–159
7. Wilkins DJ (2002) The Bathtub curve and product failure behavior. Part two- normal life and wear out. Reliability Hotwire Issue 22. http://www.weibull.com/hotwire/issue22/hottopics22.htm. Accessed Jul 10 2013

Effect of High Speed Rail Transit and Impact Loads on Ballast Degradation

Nicholas Keeng, Jun Li and Hong Hao

Abstract The emerging need of railway as a principal means of massive transport has encouraged the development of high speed trains in Australia. Large and frequent cyclic loading from heavy and fast trains leads to a progressive deterioration of the underlying railway structural system. The lack of research on degradation of ballast to counter the effects of high speed trains threatens the reliability and safety of train services and hence leads to more frequent and costly maintenance. Compaction testing with the Amsler equipment was conducted to deliver a graphical representation of fouling rates and the loading at which ballast becomes ineffective. The hammer drop test was employed to predict the service life of ballast under cyclic loading. Finite element analysis of a railway structural system subjected to a moving wheel with varying train speeds was conducted to obtain impact forces on sleeper and ballast under wheel flat effect. The deformation and stress behaviour of rail and ballast were investigated. It has been found that trains exceeding 210 km/h with a 100 mm wheel flat defect pose an immediate threat of accelerated fouling of ballast. Key findings also include the detection of different stages of ballast interaction, the observation of critical fouling force and the service life prediction of ballast under different train speeds.

N. Keeng · H. Hao
School of Civil and Resource Engineering, University of Western Australia,
Crawley, WA 6009, Australia
e-mail: 20357985@student.uwa.edu.au

H. Hao
e-mail: hong.hao@curtin.edu.au

J. Li (✉)
Department of Civil Engineering, Curtin University, Kent Street,
Bentley, WA 6102, Australia
e-mail: junli@curtin.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_45

1 Introduction

Australia has introduced faster and heavier trains in recent years due to the growing demand. This often leads to progressive track deterioration on rail and ballast of the railway system. The excessive deformation and degradation of the ballast necessitate frequent and costly track maintenance. Given the complexities of the composite track system consisting of rail, sleeper, ballast, capping and subgrade subjected to cyclic loading, current standards and design practices may be simplified for high speed applications. This research is mainly related to Western Australia, where the local rail authority has plans to introduce faster trains and increase line speeds by upgrading existing infrastructure to support growing industry and public transportation demands. In order to analyze the safety and reliability of high speed train operations and minimize maintenance costs, this paper investigates the effect of ballast degradation under cyclic loading, and impact forces due to wheel flat.

Figure 1 shows a typical railway track structure. The design of track structure needs to consider the deterioration of ballast due to breakage and subsequent implications on track deformations. Based on previous assessment of ballast characteristics [1], ballast porosity is found to be around 35–50 % and hence fouling does not become significant until the fine accumulates to 10 % or more. It is possible for the train to derail at a high speed as well as to rapidly accelerate the degradation of track structure [2]. A larger number of load cycles generally introduce the fatigue damage and increase the settlement and deformation of the ballast. Ionescu et al. [3] reported that the deformation behaviour of ballast is highly nonlinear under cyclic loading. The track structure is also often subjected to impact loads due to defects and irregularities in the wheel or rail. The impact magnitude is very high within a short duration and usually depends on the track structure and the irregularities, such as wheel flat. Wu and Thompson [4], and Remennikov and Kaewunruen [5] have demonstrated the possible ranges that the impact force may distribute in. It is observed that the impact force is around 350 kN with a travelling speed of 80 km/h and a wheel flat on the wheel [4]. Bian et al. [6] presented that the impact force of sleepers due to a wheel flat varies nonlinearly with increasing vehicle speed, and the force monotonically increases with an increasing static wheel load.

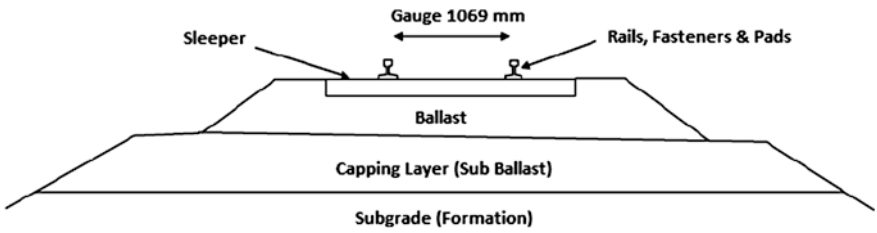


Fig. 1 A schematic description of railway track structure

2 Experimental Testing

2.1 Compaction Test

Compaction tests were conducted by using an Amsler equipment with standard ballast provided from the in-site Butler extension project in Western Australia. The supplied ballast samples from Gosnell's Quarry are considered to be consisted of 57 % granite and 43 % meta-diorite (low grade metamorphic rocks). The Amsler machine is mounted with a hydraulic jack capable of applying forces of up to 1,000 kN, and has an automatic calibration bench with the displacement and applied force sensors. Although usually used to test the strength characteristics of concrete or soil samples, the Amsler has been adapted for the purposes of this investigation with a standard 90 mm cast iron bearing plug. The compaction test investigates the behaviour when the ballast breakage under loading effect occurs and when the ballast becomes ineffective for drainage purpose as well. The reduction of voids from ballast breakage under compression will dramatically increase the rate of ballast fouling and affect track stability. A cast iron mold is filled with a series of ballast and placed on a hydraulic plate, as shown in Fig. 2a. The plate will rise with a constant displacement and a plug above the mold will crush the ballast. The compaction test was conducted with a 150 mm deep by 90 mm inner diameter mold filled with around 990 grams of ballast. Figure 2b shows the data acquisition system for the Amsler machine to measure the applied load and deformation of ballast.

When the ballast experiences compressive force, the rock fractures into smaller pieces and fills the voids between the interacting rock pieces. When the rock has been crushed to a certain degree, all the voids will have been filled up. With no available space left, the force on the rock will increase sharply and will be recorded by the sensor on the bearing plate. Figure 3a and b show the applied load and deformation measured on the ballast, respectively. It can be observed from Fig. 3b that the displacement increases linearly with a constant rate, while the applied load appears to increase significantly with the deformation as shown in Fig. 3a. At the end of the test, the load increases sharply indicating that there are no more voids to

Fig. 2 **a** Amsler equipment; **b** data acquisition system

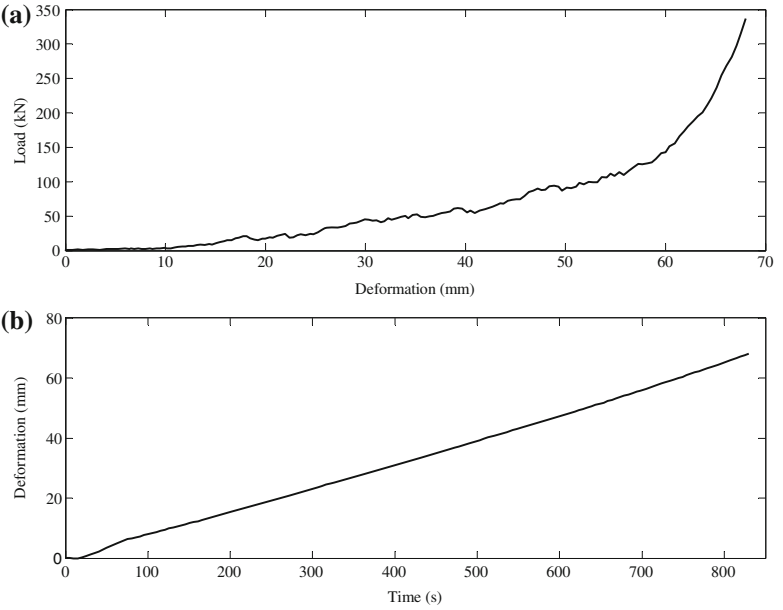


Fig. 3 Measured applied load and deformation on the ballast. **a** Applied load versus deformation, **b** Measured deformation versus time

be filled and the materials have reached their ultimate loading capacity. Three stages of ballast compression are observed. Until 50 kN, the ballast particles appear to rearrange amongst each other rather than displaying attributes of full ballast breakage. This region can be considered as the compaction region—an area that track compaction maintenance would be sought to increase the track stability. Between 50 and 150 kN, the ballast achieves a high level of frictional interlock and accelerated breakage is observed. Once the ballast has been fully compacted, the voids are no longer presented in the sample and the load line increases exponentially without any fluctuations. At 150 kN, the fine accumulation is declared to be at the maximum value. The scale factor is calculated by comparing the contact pressures between the bearing plug interface and the sleeper on ballast interface, and the critical fouling force on a track is derived as 420 kN. This can be regarded as the minimum force required to introduce significant rates of fine accumulation and accelerated ballast fouling. This is also considered as the maximum force from sleeper on ballast, which the rail can still operate with a good condition of ballast.

2.2 Hammer Drop Test

The hammer drop test is usually used in geotechnical engineering. The test utilizes a standard 4.902 kg hammer falling to impact the ballast sample with a height of 150 mm. The particle breakage of the ballast will be observed, and the service life

Fig. 4 **a** Hammer drop test; **b** sieves of varying sizes

of railway ballast under cyclic loading will be evaluated. A sample of ballast is placed in a proctor mold and a hammer weight will be repeatedly dropped as shown in Fig. 4a. Sieves of varying sizes in Fig. 4b will separate the ballast and a particle size distribution can be generated. The number of drops, degree of particle breakage can be scaled to a real size application by analyzing the correlation between force, dimensions and mass.

The initial weight distribution using sieve analysis was conducted. The hammer drop test was conducted, and the sieving of ballast was conducted at 200, 400, 600, 800, 1000 and 1500 cycles respectively to observe the ballast breakage. Four tests were conducted to get an accurate average result. Figure 5 shows the weight distribution observed with different number of hammer drop cycles. It is observed that ballast aggregates bigger than 26.5 mm get degraded faster in track conditions. This may be due to a greater number of flaws or defects found in larger aggregate as mentioned in reference [7]. Ballast retained by a 13.2 mm sieve with different cycles indicates that there is an equilibrium level where the rate of breakage above and below the sieve range is equivalent. On the other hand, particles retained between sieve sizes 6.7 and 9.5 mm do not appear to accumulate significantly.

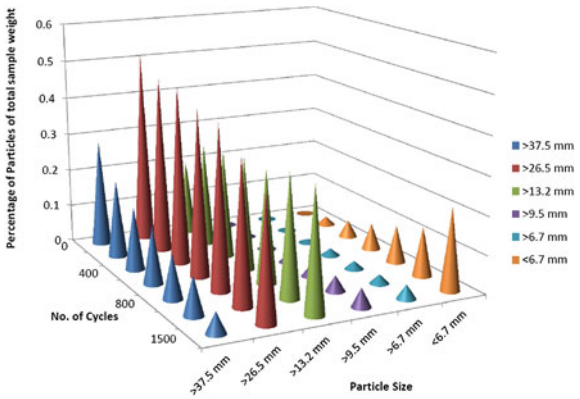


Fig. 5 Weight distribution of hamper drop test with different number of cycles

Table 1 Estimated ballast service life under varying train speeds

Train speed (km/h)	Dynamic amplification factor	Service life (years)
60	1.135	32.29–39.21
80	1.145	32.01–38.86
100	1.16	31.59–38.36
120	1.175	31.19–37.87
140	1.2	30.54–37.08
160	1.25	29.32–35.6
180	1.305	28.08–34.09
200	1.465	25.02–30.38

Particles smaller than 6.7 mm grow significantly with the growing number of cycles indicating the ballast is crushed under cyclic loading effect.

The rate of accumulation formed a linear function against the number of cycles and an equation can be formulated to determine how many cycles will produce maximal fouling material (void ratio typically varies between 0.35 and 0.425). The number of cycles was scaled against variables such as drop height, mass, and impact pressure. The ballast lifespan per wheel was then determined to be between 28.967 and 35.172 million static cycles. Based on an average of 4 carriages per train on 950 journeys per week in Western Australia lines, the ballast life is determined to be between 36.65 and 44.5 years. By considering the dynamic amplification factor, various serviceability lifespans are given for different train speeds as shown in Table 1. At higher speeds, the serviceability life decreases at an increasing rate. Beyond 200 km/h, trains encountered on the line will be heavier, and hence likely to run on a dedicated track. The track will have different design standards, and consequently, service life of ballast will subsequently change.

3 Numerical Simulation

A commercial finite element analysis package, ANSYS 14, was used to model the track structure and analyze the effect of high speed trains on ballast degradation. Structural symmetry allowed for the design of a half-track finite element model as shown in Fig. 6 with solid elements. The track structure consists of five different components, namely rail, sleeper, ballast layer, capping layer, and subgrade. The rail and sleeper components are simplified into block shapes due to the complexities in geometry. The dynamic wheel forces generated by the wheel flat are calculated as equivalent forces, and applied on the top of the rail. The transient analysis is conducted including 166 load steps with a 10 mm mesh for rail elements. The simplified track structure conforms to the same design geometry as WA's electrified urban rail system. The model consists of 62,856 nodes with selectively refined

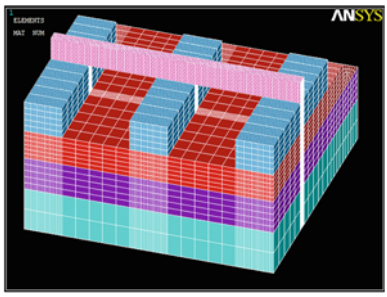


Fig. 6 Finite element model for the analysis

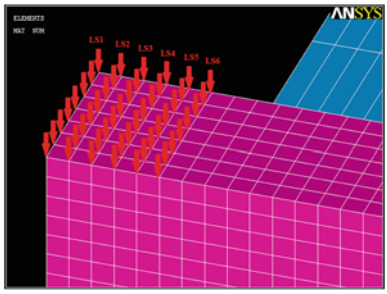


Fig. 7 Applied forces of load steps 1–6

mesh at the rail and underlying sleeper components. As the ballast degradation under the sleeper will be studied, the wheel flat impact is located on the rail above the central concrete sleeper. This location of the wheel flat would produce the maximum force encountered by the ballast.

To perform the dynamic transient analysis, load steps are defined across the length of the modelled rail to simulate the moving wheel loading. Equivalent forces are calculated and applied on the rail. The width of the rail is divided into 8 nodes with 7 divisions of 10 mm for each. Therefore the force is distributed on these 8 nodes. Each force is represented as a red arrow as shown in Fig. 7. This set of forces represents a single load step. The process is repeated along the length of the rail (excluding the location of the wheel flat). The load steps are applied one at a time with a defined time increment. There are a total of 166 load steps in the dynamic analysis. Figure 7 shows the first six load step with equivalent forces applied on the model. Modelling the wheel flat impact consists of a series of load steps being applied as a function of time at where the impact occurs. A 100 mm wheel flat is considered in this study. The forces are distributed over the nodes on the contact surface. Figure 8 shows the first four load steps simulating the equivalent forces due to wheel flat.

Maximum displacements of rail and ballast under different speeds are presented in Fig. 9. The rail displacements between 120 and 180 km/h agree well the range of displacements in a previous study [8]. The ballast layer displays very little displacement or rate of increase under the impact force and as such, it is not utilized as the primary factor to determine the safety tolerances. It is noticed that rail

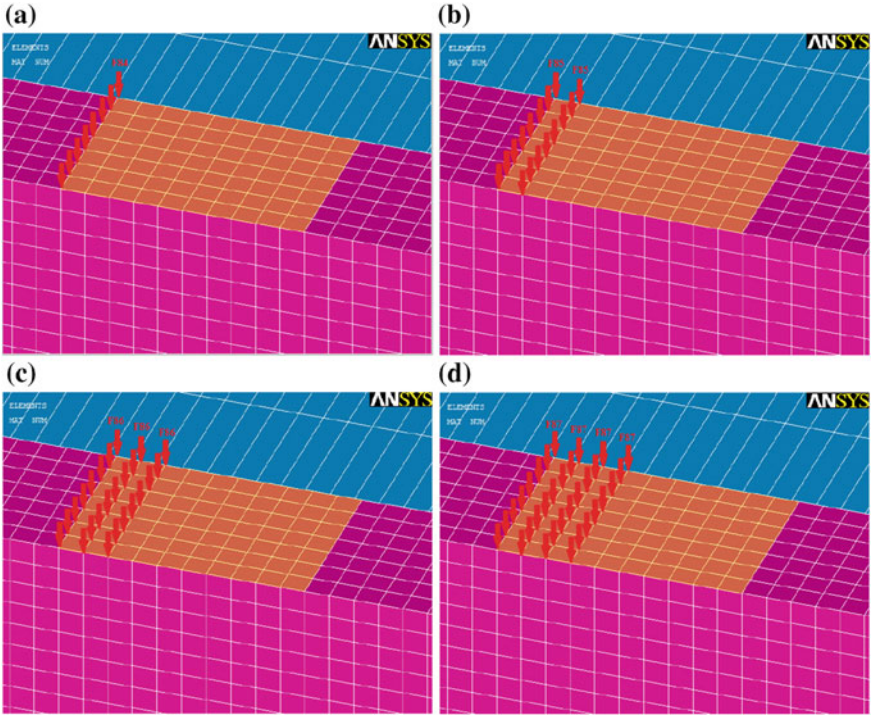


Fig. 8 Simulation of the first 4 steps with wheel flat. **a** Load step 84, b Load step 85, **c** Load step 86, **d** Load step 87

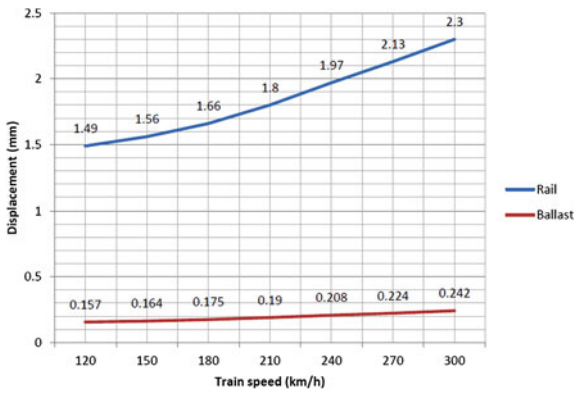


Fig. 9 Maximum displacements on rail and ballast

deflections exceeding 2 mm have been outlined as unacceptable by Public Transport Authority of Western Australia [9]. Interpolation of the rail displacements in Fig. 9 indicates that unsafe displacement (2 mm) of rail occurs at a speed of 245 km/h. Other variables outside the scope of impact forces will further encourage the construction of a higher specification track to accommodate speeds exceeding

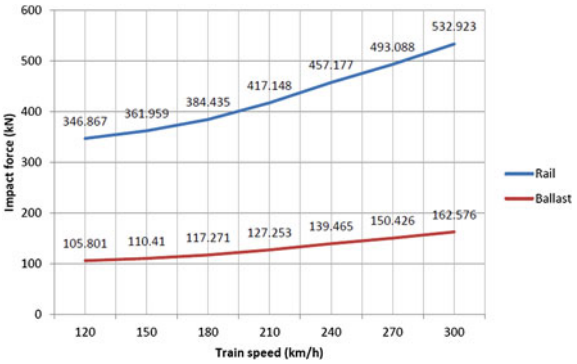


Fig. 10 Maximum impact forces on rail and ballast

245 km/h. In practice, the maximum track speed allowable before failure is never employed where safety is at the utmost concern.

Maximum impact forces are recorded at the rail-sleeper surface and the sleeper-ballast interface and they are shown in Fig. 10. The forces increase at a nonlinear rate with operating speeds. The critical fouling force of 420 kN has been observed from the Amsler test in Sect. 2. It can be derived from Fig. 10 that operating speed at 210 km/h will produce an impact force around 420 kN. Because the critical fouling force does not directly threaten the safety of passenger services, monitoring techniques can be applied. On board wheel monitoring systems are suggested for high speed trains and shall have the ability to detect wheel defects. Early detection and maintenance is vital in reducing high speed impact forces. The optimal efficiency of applying monitoring techniques, maintenance techniques and infrastructure replacement are yet to be determined and presents a new scope for future research.

It can be found from Fig. 10 that impact forces increase at a fast rate particularly beyond 180 km/h. Current mainline speeds in Western Australia do not have to worry about strong impact loads as the fastest metropolitan mainline operates at 130 km/h. In this case the rail displacement is around 1.5 mm and impact force on sleeper is 350 kN. It can be found that these match well with the reference values in a previous study [4]. It is of interest to note that the fastest trains in Australia run at 160 km/h. If the current railways are upgraded to a higher speed, with current ballast conditions, 210 km/h is a suggested maximum speed with the appropriate monitoring techniques.

Figure 11a and b show the structural displacement and the first principle stress of the track system at all load steps when the wheel travels on the rail with a speed of 210 km/h. It can be observed that the maximum displacement is observed when the wheel is located at the centre of two sleepers and the maximum stress is located at the rail-sleeper interface, which indicates the existence of impact force.

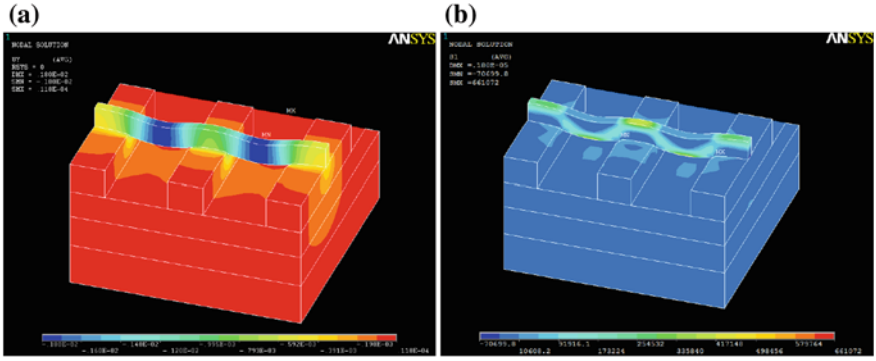


Fig. 11 **a** Structural displacement (m); **b** first principle stress

4 Conclusion

Experimental tests, namely Amsler compaction test and hammer drop test, and finite element analysis have been conducted to provide data and information in determining the effect of impact loads from high speed trains. Amsler testing presented three stages of ballast interaction under force loading, that is, re-arrangement, ballast breakage, and ineffective drainage zones. A critical fouling force of 420 kN for a narrow gauge concrete sleeper is suggested. When the force exceeds 420 kN, the ballast experiences accelerated fouling. Hammer drop tests are conducted to estimate the ballast service life, and the predicted service life of the ballast under different train speeds is given. Finite element analysis suggests that train speeds above 245 km/h exceed the general rail deformation tolerance as set out by the Code of Practice in Western Australia. Critical impact force can be observed when train speeds exceed 210 km/h with a 100 mm wheel flat. Therefore consistent wheel or track defect monitoring should be carried out in order to prevent the development of large impact forces before they pose a serious threat of accelerated ballast degradation, and guarantee a safe railway operation. With pressure to increase operating speeds and recently proposed high speed train projects in Western Australia, this research investigates several factors that shall need further considerations in order to maintain a high standard of railway transportation performance.

Acknowledgments The authors gratefully acknowledge the financial support from Australian CRC for Infrastructure and Engineering Asset Management: Project No.3104; and Australian Research Council Discovery Early Career Researcher Awards, Project No. DE140101741.

References

1. Selig ET, Waters JM (1994) Track geotechnology and substructure management. Thomas Telford, Boston
2. Gilligan A (2012) High Speed Rail Link 'at risk of derailment' because of 225 mph trains. http://www.telegraph.co.uk/news/uknews/road-and-rail-transport/9090727/High-speed-rail-link-at-risk-of-derailment-because-of-225mph-trains.html
3. Ionescu D, Indraratna B, Christie HD (1998) Behaviour of railway ballast under dynamic loads. In: 13th Southeast Asian geotechnical conference, Taipei, Taiwan
4. Wu T, Thompson D (2001) A hybrid model for wheel/track dynamic interaction and noise generation due to wheel flats. In: Institute of sound and vibration research (ISVR) technical memorandum No. 859, January 2001
5. Remennikov AM, Kaewunruen S (2007) Resistance of railway concrete sleepers to impact loading. In: 7th international conference on shock and impact loads on structures, Beijing, China
6. Bian J, Gu YT, Murray M (2013) Numerical study of impact forces on railway sleepers under wheel flat. Adv Struct Eng 16(1):127–134. doi:10.1260/1369-4332.16.1.127
7. Lade PV, Yamamuro JA, Bopp PA (1996) Significance of particle crushing in granular materials. Geotech Eng 122(4):309–316. doi:10.1061/(ASCE)0733-9410(1996)122, 4(309)
8. Banimahd M, Woodward PK (2007) 3-Dimensional finite element modelling of railway transitions. In: 9th international conference on railway engineering, Xitrack, London
9. Public Transport Authority (2011) Code of practice for the PTA narrow gauge mainline. Government of Western Australia, Perth, Western Australia, Australia

Integrating Real-Time Monitoring and Asset Health Prediction for Power Transformer Intelligent Maintenance and Decision Support

Amy J.C. Trappey, Charles V. Trappey, Lin Ma and Jimmy C.M. Chang

Abstract Large sized transformers are an important part of global power systems and industrial infrastructures. An unexpected failure of a power transformer can cause severe production damage and significant loss throughput the power grid. In order to prevent power facilities from malfunctions and breakdowns, the development of real-time monitoring and health prediction tools are of great interests to both researchers and practitioners. An advanced monitoring tool performs real-time monitoring of key parameters to detect signals of potential failure through data mining techniques and prediction models. Asset managers use the result to develop a suitable maintenance and repair strategy for failure prevention. Principal component analysis (PCA) and back-propagation artificial neural network (BP-ANN) are the algorithms adopted in the research. This chapter utilizes industrial power transformers' historical data from Taiwan and Australia to train and test the failure prediction models and to verify the proposed methodology. First, PCA detects the conditions of transformers by identifying the state of dissolved gasses. Then, the BP-ANN health prediction model is trained using the key factor values. The integrated engineering asset management database includes nine gases in oil as input factors (N_2, O_2, CO_2, CO, H_2, CH_4, C_2H_4, C_2H_6, and C_2H_2). After applying the principal components algorithm, the research identifies five factors from the Taiwan operational transformer data and six factors from the Australia data. The integrated PCA and BP-ANN fault diagnosis system yields effective and accurate predictions

A.J.C. Trappey (✉) · J.C.M. Chang
Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan
e-mail: trappey@ie.nthu.edu.tw

J.C.M. Chang
e-mail: s100034550@m100.nthu.edu.tw

C.V. Trappey
Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan
e-mail: trappey@faculty.nctu.edu.tw

L. Ma
Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia
e-mail: l.ma@qut.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_46

when tested using Taiwan and Australia data. The accuracy rates are much higher (i.e., 92 and 96 % respectively) when compared to previous result of 69 and 75 %. This research is benchmarked against the DGA heuristic approaches including IEEE's Doernenburg and Rogers and IEC's Duval Triangle for the experimental fault diagnoses.

Keywords Engineering asset management · Back-propagation artificial neural network · Principal component analysis · Intelligent prognosis · Gases in oil

1 Introduction

Transformers are a critical part of electricity transmission and distribution grid and effect the stable operation of networked components. Many chemicals are generated inside the transformer when it is operating. Bhalla et al. [2] specify the gases that effect daily working of condition of transformers include the insulating oil, hydrogen (H_2), carbon monoxide (CO), carbon dioxide (CO_2), nitrogen (N_2), oxygen (O_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4) and ethane (C_2H_6). Increases in the amount of gases within the power transformer tend to cause internal electrical or thermal failure. Electrical failures often result from arcing discharges and partial discharge which ignite gases. Thermal failure leads to a low, medium, or high temperature fault. The transformer internal fault type causes varying levels of damage to the transformers solid insulation bushings, on-load tap changer, cables and core.

There are two well known maintenance procedures which are time-based maintenance and condition-based maintenance. Time-based maintenance performs the inspection over a constant time interval using a time schedule consistent with company strategic planning. Condition-based maintenance provides a planned maintenance strategy with other benefits, as indicated by Roberts et al. [14]: (1) Determine a better time to repair or perform the maintenance activities. (2) Due to the additional information, reduce the average mean time to repair. (3) When the initial failures are found, the equipment managers have more time to develop appropriate maintenance plans. (4) Reduce the parts used for replacement.

The time-based strategy fails when no one detects the fault in between the planned maintenances. Abu-Elanien et al. [1] indicate that if the interval time is too short then unnecessary inspections will waste time and money. Therefore, efficient and effective fault detection technology and accurate detection times are important for equipment managers to manage transformer conditions. BP-ANN provides the advantage of quickly learning and adapting to the data sets consisting of the transformers' condition and the corresponding fault signals without needing to know the relationship between data conditions and signals. Furthermore, the algorithm does not need experts to identify oil sample results. In this chapter, we propose an intelligent on-line diagnostic system based on the PCA and BP-ANN to

improve the condition-based asset management. A systematic and generic approach is used to manage the on-line diagnostic tools which monitor, access information, and recommend maintenance strategies.

2 Literature Review

The concept of engineering asset management has developed for decades, and has been applied in asset management of various industries. This research focuses on the transformer fault diagnosis with the objective to increase the efficiency of asset management. In this section, we focus on engineering assets management, principal component analysis and back-propagation neural network fault diagnosis.

2.1 Engineering Asset Management

The early concept of Engineering Asset Management (EAM) was proposed by Parkes [12]. The UK standard PAS 55-1 [13] is commonly used to define EAM. The purpose of EAM is to develop an optimization strategy of asset management regarding performances, risks and expenses in the overall life cycle of assets. Hastings [8] also regarded EAM as a series of activities to help organizations or enterprises achieve their goals. For many organizations, EAM has become an important part of daily management, especially for the investment of equipment and infrastructure. Ma [11] states that the current approaches are not sufficient for EAM and requires new techniques and models that are more reliable and robust for practicing the state of physical assets. Through monitoring and analysis of asset behavior, the underlying symptoms of fault can be detected with greater efficiency. Ma et al. [10] proposed an agent-based asset management platform which is used in communication, coordination and allocation between the power assets to maintain the electric facilities. Trappey et al. [18] proposed a multi-agent system to develop the maintenance decision supporting system for the large size transformers. The system embeds a negotiation mechanism to strengthen the cooperation of the entire collaborative maintenance chain. El-Hag et al. [5] extracted acoustic features and utilized radio frequency to detect partial discharge signals so as to evaluate the health state of insulation oil.

2.2 Principal Component Analysis for Extracting Engineering Parameter

Principal component analysis (PCA) is a multivariate statistical method proposed by Pearson in 1901 and it was further developed by Hotelling and became a

statistical method in 1933. Therefore, PCA is also known as Hotelling transform. The main purpose of PCA is to find an orthogonal transformation matrix and to project data into a new axis to reduce the dimension of variables through a linear combination of multiple variables. Since the aim of PCA converts original variables to comprehensive new indicators, it has the following advantages. The approach retains most of the original variable information but also decreases the complexity of the data. For the air-handling processes air conditioning systems, Sakthivel et al. [21] enhanced the PCA-based method in fault diagnosis. They presented an expert-based multivariate decoupling method which can identify unique fault patterns of sensors by analyzing the physical cause-effect relations among variables. For the mono-block centrifugal pumps, Elangovan et al. [16] used PCA-based decision tree-fuzzy sets to detect faults and compared fuzzy decision trees and rough sets to derive a more effective method. Elangovan et al. [4] analyzed cutting tool vibration signals to discover the available hidden information for improving machining activity and processing quality. They reduced variable dimensions from 12 to 6 using PCA and raised the fault diagnosis accuracy from 73 to 87 %.

2.3 Back-Propagation Artificial Neural Network for Engineering Asset Management

The Back-Propagation Artificial Neural Network (BP-ANN) is the most representative model of the learning neural network and is commonly applied to various fields. It belongs to back-forward network architecture and has a supervised learning process which is suitable for applications of diagnosis, prognosis, and forecasting. Werbos [19] first added hidden layers in the neural network. In addition, Rumelhart et al. [15] proposed a general rule and published a back-forward artificial neural network to further affirm the value of BP-ANN. So far, it is still one of the most useful neural network approaches which possess the advantages of high-level learning, quick recall speed, and allowing output with consecutive values. Furthermore, BP-ANN is able to process complex samples and nonlinear problems and it can be widely applied in various fields. Shintemirov et al. [17] applied a genetic algorithm to extract the key parameters of the transformers as input factors of ANN, support vector machine, and the k-nearest neighbor approach to diagnose fault types. Ghunem et al. [6] measured the furan content to predict the aging state of oil impregnated insulators in the transformer by ANN. Bhalla et al. [2] extracted the relationship between input layer and output layer in ANN as a rule-based diagnosis method to detect incipient faults.

3 Methodology

3.1 Principal Component Analysis

Before applying PCA, the correlation of each variable should be measured by Kaiser-Meyer-Olkin (KMO) analysis [20]. The equation of KMO is shown in Eq. 1.

$$KMO = \frac{\sum_{i} \sum_{j(i-j)} r_{ij}^2}{\sum_{i} \sum_{j(i-j)} r_{ij}^2 + \sum_{i} \sum_{j(i-j)} s_{ij}^2}. \tag{1}$$

In Eq. 1, r_{ij} is the correlation coefficient of x_i and x_j, and s_{ij} is the partial correlation coefficient of x_i and x_j. If the KMO value is closer to 1, there is a high level of correlation and it is more suitable to use PCA. However, if KMO value is less than 0.5, it is not suitable to use PCA.

PCA is one of methods which can effectively explain the variability from raw data and simplify it into fewer critical variables from multiple variables using the linear combination. At the same time, it retains the characteristics and information of the original variables. The PCA process is shown below.

Step 1. According to data define the functioned relationship between components and variables.

Step 2. In this research, the measurement units are the same so that the sample variance indicates that it is more appropriate to use the covariance matrix (S) in PCA:

Step 3. In order to obtain maximum variation, the restriction $a' \cdot a = 1$ is applied and the maximization of $a' \cdot S \cdot a$ follows:

$$\begin{aligned} &Max\, Var(Z) = a' \cdot S \cdot a \\ &ST \quad a' \cdot a = 1 \end{aligned} \tag{2}$$

Step 4. According to the Lagrange formula:

$$L = a' \cdot S \cdot a - \lambda(a'a - 1) \tag{3}$$

Using the Eq. (3) with partial differentiation and setting the equation to 0.

$$(S - \lambda I) \cdot a = 0 \tag{4}$$

Since Eq. (4) has an infinite number of solutions, the restriction $|S - \lambda I| \cdot a = 0$ is included.

Table 1 The relationship between sample size and level of significance

Sample size	350	250	200	150	120	100	85	70	60	50
Level of significance	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75

Step 5. Solving Eq. (4), the feature vector a and corresponding eigenvalues are computed.

$$\lambda_i = a' \cdot S \cdot a = s_{Yi}^2 \tag{5}$$

The total variance is represented by:

$$\lambda_1 + \lambda_2 + \lambda_3 + \cdots + \lambda_p = trS \tag{6}$$

Step 6. Through repeated iterations, the explanatory power of PCA and factor loading are obtained. Given the factor loading in the each component, the degree of importance between sample size and the level of significance is determined [7] as shown in Table 1.

3.2 Back-Propagation Artificial Neural Network

Gas in oil is released by the degradation of transformer oil and may be used to predict the failure of the device. PCA defines the value for the neural network nodes which are normalized as binary valuables. The elements for the BP-ANN model are the nodes of the input layer, the nodes of the hidden layer, the number of hidden layers, the learning rate, and the number of iterations. The BP-ANN model can use multiple hidden layers, but excessive hidden layers increases the complexity of the network and decrease the convergence rate, which decreases the learning of the network. Therefore, a single hidden layer increases accuracy [3, 9]. The following equation is used to set the number of nodes for the hidden layers: $H = \frac{1}{2} \times I \times O$ and $H = \frac{1}{2} \times (I + O)$. Where H is the hidden layer nodes, I is the input layer nodes and O is the output layer nodes.

In this chapter, we use one hidden layer to construct the framework of BP-ANN. The formula for computing weights is defined by:

$$H_j = f\left(net_j\right) = f\left(\sum_i \left(w_{ij} \cdot x_i\right)\right) \tag{7}$$

where x_i is the value of input factor and w_{ij} is the corresponding weight. The output value of the hidden layer H_j is calculated by the activation function using the sum of input values multiplied by the corresponding weight.

$$f(x) = \frac{1}{(1 + e^{-x})} \tag{8}$$

The preprocessed data is converted to binary for transformer diagnosis and a sigmoid function is used as the activation function of each layer.

The output value of BP-ANN is calculated as the hidden layer:

$$O_k = f(net_k) = f\left(\sum_j w_{jk} \cdot H_j\right) = f\left(\sum_j w_{jk} \cdot f\left(\sum_i \left(w_{ij} \cdot x_i\right)\right)\right). \tag{9}$$

Each iteration calculates an error value. The root-mean-square error represents the BP-ANN error rate.

$$\delta_k = (O_k - T_k) \cdot O_k \cdot (1 - O_k) \tag{10}$$

where, O_k is the predicted output of the network and T_k is the actual target alarm condition.

4 System Development and Implementation

The system integrates two modules for transformer fault diagnosis. One module is used for setting each user's access feature. The other module which includes KMO, PCA, and BP-ANN is used for gas numerical analysis, the extraction of key factors, and the data mining and prediction models. Transformer operation data is used for the diagnostic model. The data set consists of 260 data points from Taiwan and 240 data point from Australia. The datasets are readings from gases in oil (i.e. N_2, O_2, CO_2, CO, H_2, CH_4, C_2H_4, C_2H_6 and C_2H_2), furanic compounds (i.e. 5-HMF, 2-FAL, 2-FOL, 2-ACF and 5-MEF), and voltage levels from transformers in service. The furan contents are used to detect insulation deterioration in the transformer [1] and dissolved gases in oil are used as prediction factors. The predicted statuses (normal, waiting acknowledgement, abnormal) are binary variables.

According to the literature, oxygen (O_2) and nitrogen (N_2) are not the main factors which result in electrical stress, thermal stress and thermal insulation deterioration (IEEE C57.105, 2008; [2]). Thus, we exclude O_2 and N_2 before executing PCA and BP-ANN. Before factors screening, KMO analysis is conducted using values of 0.62 and 0.79 in Taiwan and Australia respectively. The datasets of both Taiwan and Australia are used for the principal component analysis and the cumulative variance percentage of second principal component are greater than

Table 2 Factors loading for Taiwan transformers

Variable	H_2	CO_2	C_2H_4	C_2H_6	C_2H_2	CH_4	CO
Component 1	0.0450	0.9340*	−0.3680	−0.3480	0.0170	−0.3720	0.5600*
Component 2	0.2110	0.3560	0.9200*	0.9080*	0.0580	0.9150*	0.1500

* The factor loading value is greater than the threshold value 0.4

Table 3 Factors loading for Australia transformers

Variable	H_2	CO_2	C_2H_4	C_2H_6	C_2H_2	CH_4	CO
Component 1	0.7010*	0.6730*	0.5350*	0.8130*	0.3450	0.8290*	0.8550*
Component 2	0.3580	0.3620	0.2030	0.3260	0.1450	0.3280	0.3610

* The factor loading value is greater than the threshold value 0.4

90 % for both the data sets so two principal components are used. According to Table 1, the significant factor loading value of Taiwan and Australia data have to be greater than 0.4 respectively. As shown in Table 2 and Table 3, we extract five key factors (CO_2, C_2H_4, C_2H_6, CH_4 and CO) from Taiwan data and six key factors (H_2, CO_2, C_2H_4, C_2H_6, CH_4 and CO) from Australia data.

The momentum is usually set between 0.1 and 0.8 [22]. In this research, we set it as 0.8. For the Taiwan transformer BP-ANN model, we use 206 datasets (80 % of available data points) to train the models, 28 datasets (10 % of available data points) to test the models, and 26 datasets (10 % of available data points) to verify the diagnostic BP-ANN models. Through a series of experiments, we import the data into the BP-ANN diagnostic platform in order to obtain a good diagnostic model. We obtain an optimal structure for the transformer fault diagnostic model being 5-5-3 with high accuracy. Furthermore, we use 191 datasets (80 % of available data points) to train the model, 25 datasets (10 % of available data points) to test the model, and 24 datasets (10 % of available data points) to verify the diagnostic BP-ANN model for Australia case. Through a series of experiments, we obtain an optimal structure for transformer fault diagnostic model, 6-3-6, which also yields high accuracy (Table 4).

Table 4 The experimental result of Australia transformer diagnostic model construction

No.	Input/output layer nodes	Hidden layer nodes	Momentum	Final learning rate	Accuracy rate	Error rate
1	6/3	3	0.8	0.8866	0.8981	0.1019
2	6/3	4	0.8	0.8824	0.8904	0.1096
3	6/3	5	0.8	0.6217	0.9231	0.0769
4*	6/3	6	0.8	0.8799	0.9738	0.0262
5	6/3	7	0.8	0.7989	0.9698	0.0302
6	6/3	8	0.8	0.8576	0.9691	0.0309
7	6/3	9	0.8	0.8812	0.9564	0.0436
8	6/3	10	0.8	0.8941	0.9233	0.0767
9	6/3	11	0.8	0.8723	0.9023	0.0977

* The best experiment result

Table 5 Accuracies of diagnostic results using Taiwan and Australia transformer data

	Actual data	TW-C_1	TW-C_2	TW-C_3	AUS-C_1	AUS-C_2	AUS-C_3
Predicted data	C1	14	0	0	14	0	0
	C2	1	5	0	0	6	1
	C3	0	1	5	0	0	3
	Summation	15	6	5	14	6	4
	Overall accuracy		Taiwan 92%			Australia 96%	

*C_1 - normal; C_2 - waiting acknowledgement; C_3 - abnormal

Through multiple experiments, the optimal BP-ANN model structure is established. In these experiments, we import real data into the system for verification and identify the status of the transformers. Table 5 shows verified results of the optimal fault diagnostic model derived from PCA and BP-ANN. The accuracy of 92 and 96 % are achieved for the Taiwan and Australia transformer's datasets. However, poorer results of 69 and 75 % are achieved when the simple BP–ANN without PCA is applied for prognosis. If we import the real transformer data into the diagnostic system to build BP-ANN models without using PCA, the result of classification will poorly predict actual transformer conditions. This increases the probability of misjudgment since the mapping relationship between the input and output layer is difficult to express.

5 Conclusion

This chapter proposes a fault diagnostic decision support system for power transformer asset maintenance and management. Transformer data from Taiwan and Australia are used to train fault diagnostic models. Through related literature review and a series of experiments, we extract key factors from the transformer data and improve accuracy of classification. Furthermore, we know that the high-dimension original data results in high correlation between each variable and causes poor prediction results. If we utilize PCA to obtain key factors, it decreases the relationship of each variable and we use low-dimension variables to express high-dimension variance. In summary, the equipment managers can effectively grasp the operational status of power transformer and obtain valuable information for decision support and maintenance. Through the development of diagnostic system and extraction of engineering parameters, unexpected equipment damage and unnecessary losses can be prevented.

References

1. Abu-Elanien AEB, Salama MMA (2010) Asset management techniques for transformers. Electr Power Syst Res 80(4):456–464
2. Bhalla D, Bansal RK, HiO Gupta (2012) Function analysis based rule extraction from artificial neural networks for transformer incipient fault diagnosis. Electr Power Energy Syst 43 (1):1196–1203
3. Cybenko G (1989) Approximation by superpositions of a sigmoidal function. Math Control Signals Syst 5(4):303–314
4. Elangovan M, Babu Devasenapati S, Sakthivel NR, Ramachandran KI (2011) Evaluation of expert system for condition monitoring of a single point cutting tool using principle component analysis and decision tree algorithm. Expert Syst Appl 38(4):4450–4459
5. El-Hag AH, Saker YA, Shurrab IY (2011) Online oil condition monitoring using a partial-discharge signal. IEEE Trans Power Deliv 26(2):1288–1289

6. Ghunem RA, El-Hag AH, Assaleh K (2010) Prediction of furan content in transformer oil using artificial neural networks (ANN). In: IEEE international symposium on electrical insulation (ISEI), San Diego, CA, USA, June 6–9, 1–4
7. Hair J, Anderson R, Tathan R, Black W (1998) Multivar data anal. Macmillan, NJ
8. Hastings NAJ (2010) Physical asset management. Springer, London
9. Hornik K, Stinchcombe M, White H (1989) Multi-layer feedforward networks are universal approximations. Neural Netw 2(5):336–359
10. Ma C, Tang WHT, Yang Z, Wu QH, Fitch J (2007) Asset managing the power dilemma. IEEE Control Autom Mag 18(5):40–45. IEEE Press, October–November
11. Ma L (2007) Condition monitoring in engineering asset management. Asia-Pacific Vibration Conference (APVC), August 6–9, Sapporo, Japan, pp 1–16
12. Parkes D (1978) Terotechnology handbook. Her Majesty's Stationery Office, London
13. PAS 55-1 (2008) Asset management: specification for the optimized management of physical assets. British Standards Institution, UK
14. Roberts C, Dassanayake HPB, Lehrasab N, Goodman CJ (2002) Distributed quantitative and qualitative fault diagnosis: railway junction case study. Control Eng Pract 10(4):419–429
15. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. Parallel Distrib Process: Explor Microstruct Cognit 1:318–363
16. Sakthivel NR, Sugumaranb V, Nair BB (2010) Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump. Mech Syst Signal Process 24(6):1887–1906
17. Shintemirov A, Tang W, Wu QH (2009) Power transformer fault classification based on dissolved gas analysis by implementing bootstrap and genetic programming. IEEE Trans Syst Man Cybern-Part C: Appl Rev 39(1):69–79
18. Trappey AJC, Trappey CV, Ni WC (2013) A multi-agent collaborative maintenance platform applying game theory negotiation strategies. J Intell Manuf 24(3):613–623
19. Werbos P (1974) The roots of backpropagation. Wiley, Canada
20. Wu ML (1999) SPSS & the application and analysis of statistics. Wu Nan Publishing Company, Taiwan
21. Xiao F, Wang SW, Xu XH, Ge G (2009) An isolation enhanced PCA method with expert-based multivariate decoupling for sensor FDD in air-conditioning systems. Appl Therm Eng 29(4):712–722
22. Zurada JM (1992) Introduction to artificial neural systems. West Publishing Company, Minnesota

Bridge Deterioration Modeling by Markov Chain Monte Carlo (MCMC) Simulation Method

N.K. Walgama Wellalage, Tieling Zhang, Richard Dwight and Khaled El-Akruti

Abstract There are over 10,000 rail bridges in Australia that were made of different materials and constructed at different years. Managing thousands of bridges has become a real challenge for rail bridge engineers without having a systematic approach for decision making. Developing best suitable deterioration models is essential in order to implement a comprehensive Bridge Management System (BMS). In State Based Markov Deterioration (SBMD) modeling, the main task is to estimate Transition Probability Matrixes (TPMs). In this study, Markov Chain Monte Carlo (MCMC) simulation method is utilized to estimate TPMs of railway bridge elements by overcoming some limitations of conventional and nonlinear optimization-based TPM estimation methods. The bridge inventory data over 15 years of 1,000 Australian railway bridges were reviewed and contribution factors for railway bridge deterioration were identified. MCMC simulation models were applied at bridge network level. Results show that TPMs corresponding to critical bridge elements can be obtained by Metropolis-Hasting Algorithm (MHA) coded in MATLAB program until it converges to stationary transition probability distributions. The predicted condition state distributions of selected bridge element group were tested by statistical hypothesis tests to validate the suitability of bridge deterioration models developed.

N.K. Walgama Wellalage · T. Zhang (✉) · R. Dwight · K. El-Akruti
University of Wollongong, Wollongong, NSW 2522, Australia
e-mail: tieling@uow.edu.au

N.K. Walgama Wellalage
e-mail: wwnk807@uowmail.edu.au

R. Dwight
e-mail: radwight@uow.edu.au

K. El-Akruti
e-mail: khaled@uow.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_47

1 Introduction and Background

Bridge inspection data consist of condition ratings of main components such as superstructure, deck, substructure, etc., or sub key components. Although the deterioration processes of bridge components are continuous, discrete condition ratings are used to measure the level of deterioration of components to reduce the complexity of the continuous condition monitoring [1]. Ratings are usually assigned on different scales by different organizations and, inspections are normally conducted once in every year or two years. For an example, Federal Highway Administration (FHWA) in USA uses range from 0 to 9 whereas railway bridge organizations in Australia assigns ratings on a scale of 1–6 or 1–4. If reliable bridge condition rating data are available for relatively long period of time, that can be used to develop bridge component deterioration models [1–3]. Bridge deterioration models are used to predict the future condition states of bridge components/bridges and those are essential components of any promising Bridge Management System (BMS).

There are approximately 15,000 bridges in Australia's rail network. These bridges are made of different materials, constructed at different years. Furthermore, they are subjected to different magnitudes, frequencies and distribution of rail loading and exposed to different environmental categories; inspected and maintained by separate organizations with various inspection and maintenance standards. These uncertainties emphasize the need of probabilistic deterioration models over deterministic approaches. According to Nielsen et al. [4], any of the current inspection and maintenance practices within the Australian rail bridge industry doesn't have capability to predict the future conditions of bridge components and Australia's rail bridges seem lack of historical inspection data. Decision making procedure is subjective and it doesn't optimize the cost. Dealing with thousands of bridges has become a real challenge for bridge engineers and managing of these structures is extremely difficult without having a systematic way for decision making. Currently projects are undergoing to implement a bridge maintenance system (BMS) for rail bridges in Australia and thus best suitable deterioration models which match with current inspection and maintenance regimes are vital parts for solving the above discussed issues. This is the motivation of this study.

Markov chain approach is the most popular network level stochastic deterioration modeling technique that has been intensively used for predicting the future conditions of network level infrastructure facilities [3]. It uses available current condition rating data for predicting the future condition states while capturing the physical, model and statistical uncertainties [4]. Since Australia's rail bridge network lacks historical data, Markov approach is more suitable for developing network level deterioration models compared to neural network and risk based models. Markov models can be subdivided into state-based models and time-based models. Due to high variability of field data collected and current maintenance records of the condition state of bridge components over constant inspection period, use of discrete time sate-based deterioration models are more realistic than time-based one [5].

Therefore, discrete state Markov models are selected to establish deterioration models in this study. Main task of the Markov model here is to estimate the transition probability matrix with limited inspection data, which is also known as calibration of Markov models [6]. If Transition Probability (TP) and initial condition are known for a given component group, the future condition states can be easily obtained by using Markov chain method.

Most widely used Markov-model calibration technique is regression based nonlinear optimization approach [5]. In this method, the bridge performance curve is first obtained using linear or nonlinear regression analysis for assumed function type, normally a 3rd order polynomial [2, 5]. Secondly, constrained nonlinear optimization method is applied to minimize the sum of absolute distances between regression performance function values and related expected performance function values obtained by Markov formula which is also known as minimizing the objective function. Finally, main elements in transition probability matrix are supposed to be obtained at global minimum point of nonlinear objective function. Methodology related to distribution based nonlinear optimization is also quite similar to regression based nonlinear optimization, but the only difference is objective function that is used for the analysis, which is the sum of absolute differences between the distribution of condition obtained from the field data and the distribution given in the condition state vector from Markov equation. However, both of the nonlinear optimization methods discussed above have some common drawbacks: (1) It may stop at local minimum points resulting in incorrect transition probability values; (2) it cannot provide confidence limits of the transition probabilities and (3) it is difficult to update when new data are available [6]. Furthermore, accuracy of the regression based nonlinear optimization method is solely dependent on assumed function type that is selected for the regression fitting. Micevski [3] and Tran [6] have successfully used Markov chain Monte Carlo (MCMC) method for pipe deterioration modeling by overcoming the above mentioned limitations. Therefore, in this study, MCMC is applied to railway bridge deterioration modeling.

2 Markov Approach to Bridge Deterioration Modeling

2.1 Factor Identification

Bridge components deteriorate with time and deterioration rates and patterns may vary with contributing factors such as age, rail-traffic volume (Tonnage passes on bridge for given time), span, number of tracks, material type, functional classification (passenger train bridges or freight train bridges), nature of the defect, structure type and environmental categories, etc. These factors were identified based on most common contributing factors that were considered in previous studies [1, 2, 5] and through considering expert opinions of rail bridge engineers.

2.2 Markov Approach

The Markov chain is a special case of the Markov process and generally a discrete-time stochastic process $\{X_{(t)}, t = 0, 1, 2, \ldots\}$ that takes on a finite or countable number of possible discrete states. This can be modeled as a serious of transitions between certain states. For example, according to condition rating system of a large railway company in Australia, condition state of a bridge component can be defined by an integer between 1 and 5, where 1 represents the structure is in its best condition possible and 5 represents the maximum condition state before a bridge/component is repaired or replaced. It is assumed that whenever the process is in certain state i, there is a fixed conditional probability p_{ij} that a component will be in state j in one time unit later and can be expressed as discrete parameter stochastic process as given in Eq. (1.1). In homogenous Markov process, it is assumed that the conditional probability does not change over given time. Therefore, Eq. (1.1) can be deduced to Eq. (1.2) with all 5 states of i and j for all t's.

$$p_{ij} = \mathrm{P}\left\{X_{t+1} = j | X_{(t)} = i, X_{(t-1)} = i_{t-1}, \ldots, X_1 = i_1, X_{(0)} = i_0\right\}, \tag{1.1}$$

$$p_{ij} = \mathrm{P}\left\{X_{t+1} = j | X_{(t)} = i\right\}. \tag{1.2}$$

These probabilities are represented in matrix form that is called Transition Probability Matrix (TPM or P) of the Markov chain. For example, according to five possible condition states, it yields a 5 × 5 matrix as given bellow.

$$\mathrm{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} \end{bmatrix}. \tag{1.3}$$

Each element in the TPM represents the probability of transition from one state to another for one inspection period. Sum of the each raw of the TPM is equal to one from total probability theorem. Without rehabilitation or repair work, bridge components would be gradually deteriorating and thus, the bridge component condition ratings are either increased to a higher number or remain unchanged in one inspection period [2]. Hence, the probability p_{ij} is null for $i > j$ where i and j are condition states of the Markov model. Furthermore, in many studies [2, 5], it was assumed that bridge component condition rating would not be increased by more than one state within one single year, or probability of deteriorating to more than one state within one year is assumed to be zero. Therefore, one year TPM can be simplified further.

However, in two-year TPM, multi state transition events are automatically generated according to Markov property. This two-year TPM is equal to the second power of one year TPM [5].

Percentage of bridges/components in each condition rating after t years of age for a selected group can be expressed as a raw vector which is defined as the condition state vector ($C_{(t)}$), and expressed as $C_{(t)} = [C_{1(t)}\ C_{2(t)}\ C_{3(t)}\ C_{4(t)}\ C_{5(t)}]$, where $C_{i(t)}$ is percentage of bridge components in condition rating i (for i = 1, 2, 3, 4, 5) after t years. Furthermore, condition state matrix after 0 years (when $t = 0$) is known as the initial condition state vector which is notated as $C_{(0)}$. If initial time is chosen as age = 0, just after construction of component/(s), it is obvious that all components are in condition rating 1 (best condition). Thus, $C_{(0)} = [1\ \ 0\ \ 0\ \ 0\ \ 0]$. If transition probability matrix (TPM) and initial condition state matrix ($C_{(0)}$) are known, condition state matrix after time t can be obtained by the multiplication of initial condition state matrix by t^{th} power of TPM by using Chapman-Kolmogorov formula as follows.

$$\mathbf{C}_{(t)} = \mathbf{C}_{(0)} \times \mathbf{P}^{t} \tag{1.4}$$

Since $C_{(0)}$ is frequently known parameter for determining the future condition states, the real challenge is to estimate the Transition Probability Matrix (TPM) for a given component group. Different statistical methods have been applied in past studies to estimating TPM of infrastructure facilities including bridges, pipe lines, pavement systems, etc. Estimating of TPMs is also known as calibrating the Markov chains [6] and discussed in next section.

Performance index is defined by PI = 6 − CR where CR is the condition rating. Without a repair or replacement of a given component, PI decreases as the component age increases.

If the condition state matrix after time t is known, expected value of facility condition at time t can be calculated and defined as Expected Performance Index (EPI) [6].

$$\mathrm{EPI}_{(t)} = \mathbf{C}_{(t)} \times \mathbf{S} \tag{1.5}$$

where, $\mathbf{C}_{(t)}$ is condition state vector given in Eq. (1.4) and $\mathbf{S}$ is a column vector with condition ratings that is the transpose of matrix of [5 4 3 2 1].

3 Calibrating the Markov Model by MCMC

3.1 Bayesian Approach

Let us consider a set of data (condition ratings) for a bridge element group as $Y = \{y_1, y_2, y_3, \ldots, y_n\}$ and θ represent unknown model parameter vector (in here unknown elements p_{ij} in TPM). The joint probability distribution $P(Y/\theta)$ is known as the sampling distribution or likelihood function which should be a known

parameter to perform any inference. P(θ/Y) is known as the posterior distribution or target distribution and P(θ) is called prior distribution of unknown model parameter. According to Bayes' rule for known value of data Y, posterior distribution of model parameter is given by:

$$\mathrm{P}(\theta/\mathrm{Y}) \propto \mathrm{P}(\theta)\mathrm{P}(\mathrm{Y}/\theta). \tag{1.6}$$

When it applies to Markov calibrating model, likelihood function of unknown transition probability density vector for a given bridge data set $\mathrm{Y} = \{\mathrm{y}_1, \mathrm{y}_2, \mathrm{y}_3, \ldots, \mathrm{y}_\mathrm{n}\}$ can be easily derived by using joint probability theory and deduced into logarithmic form for easy computation [6] as:

$$\mathrm{Log}[\mathrm{L}(\mathrm{Y}/\theta)] = \sum_{t=1}^{T}\sum_{i=1}^{5} N_i^t \log(C_{it}) \tag{1.7}$$

where L (Y/θ) is the likelihood to observe a condition rating data set $\mathrm{Y} = \{\mathrm{y}_1, \mathrm{y}_2, \mathrm{y}_3, \ldots, \mathrm{y}_\mathrm{n}\}$ for given bridge element group with n total records, t is the bridge element age in years, T is the largest age found in the data set and N_i^t is the number of elements in condition i at year t; C_{it} is the probability in condition state i at year t and can be expressed as a function of TPM by using Eq. (1.4).

3.2 MCMC Simulation Method with MHA

Markov Chain Monte Carlo (MCMC) methods have been increasingly used in recent years for simulating complex, nonstandard and multivariate distributions [3]. The Metropolis-Hasting Algorithm (MHA) is the most popular example of a MCMC method and recently used for many engineering applications [6]. According to Eq. (1.6), posterior density of transition probability values is proportional to multiple of prior density into likelihood function. This property is used in this analysis to allow MCMC method with Metropolis Hasting Algorithm (MHA) to generate samples from posterior distribution [6]. When applying MHA, it is required to choose a proposal density q(x, y) where $\int q(x, y)dy = 1$, for sampling from the target distribution [6]. Although the proposal distribution is arbitrarily chosen from some family distributions, performances are dependent on how much the selected distribution approximates the posterior. "The candidate-generating density depends on the current state of the Markov chain, which means that when a process is at the point x this density generates a point y from q(x, y). The new point y is always accepted, otherwise, y can be accepted with a probability α(x, y). In other words, if the jump goes 'uphill', it is always accepted; if 'downhill', it is accepted with a non-zero probability" [6].

4 Case Study

According to data availability and by considering contribution factors such as material, average tonnage passes per week, environmental categories, etc., condition rating data of 40 transverse timber bridge decks in major inland railway lines, over past 15 years were selected to do this analysis.

A major problem identified from data analysis is related to subjective nature of bridge inspections. Veshosky [7] argued that condition ratings that are assigned by different inspectors for same bridge component potentially could result in different values. This problem has been addressed up to some extent by conducting workshops and training programs for inspectors, reviewing and adjusting the condition rating data by supervisor and by conducting detailed inspections by experienced engineers. Usually, all bridges in a one railway line are inspected by same inspector. Thus, consideration of condition rating data of bridges in one railway line for this analysis also helps to control subjective nature of the bridge inspection records based on assumption that inspection procedures and rating criteria are approximately same than across many railway lines. However, each railway line does not have significant number of bridges and hence this approach is unable to apply for each line. In this case, analysis has to be done by combining bridges in different railway lines with similar characteristics based on assumption that the observed bridge condition ratings are randomly distributed about their true values.

In this study, deterioration models were developed for railway bridge decks with no improvement work has been undertaken in between the two inspections. Therefore, inspection records for bridge deck element, after repair and reconstruction actions, have been removed from analysis data base. Nevertheless, it has been identified that every repair and maintenance work has not been recorded in bridge inventory. Some Bridge deck ratings figured improvement of condition with time. However, unless repair or maintenance work is done, bridge components would be gradually deteriorating so that the bridge condition rating is either unchanged or raised into a higher number according to the condition rating system. Rely on that assumption, bridge deck element whose condition rating had been improved over the years were identified and also removed from data base. Furthermore, very good condition rating values have been observed for relatively old bridge decks elements. This could be happen due to unrecorded repair or reconstruction work before 15 years back, since bridge agency has only 15 years back inspection records. Moreover, some of the newly constructed bridge components had not been assigned into condition rating in category one. Madanat [1] argued that this could be due to inadequate initial design or, construction or misapplication of rating procedure by bridge inspectors. It was further identified that double counting of same records also exists. Hence, these unusual condition rating data were filtered and rejected from the analysis data base. Finally, 242 total records were obtained after filtration process. For statistical validation and comparison

purpose, data set was split randomly such that 75 % as calibrated data set and the rest as the test data set. Calibrated data set was used for analysis first and validated with test data set by using Chi-square test [6].

4.1 Analysis Results

MATLAB programming codes were developed with MHA algorithm for MCMC. The MHA ran 50,000 iterations for the calibration data set and later, for entire data set until the transition probability values converge to stationary distributions after first 15,000 warm-up runs. Variance covariance matrix was adjusted until acceptance rate becomes near to the optimum acceptance rate of 0.234. Figure 1 shows the trace plots after 20,000 iterations with no warm-up runs for P11, P22, P33 and P44. It is clear that after 20,000 iterations, all transition probability values are approximately convergent to stationary distributions. Trace plots for up to 50,000 iterations after 15,000 burnings (Warm-up runs) are given in Fig. 2. Standard deviation for each transition probability values were found very small and given in Table 1 in Sect. 4.2. Mean values of each TPM elements are convergent to constants as shown in Fig. 3. Finally, mean values are obtained for the transition probability matrixes for one year transition period and two years transition period, respectively, see, Eqs. (1.8) and (1.9). By using estimated TPM and known initial

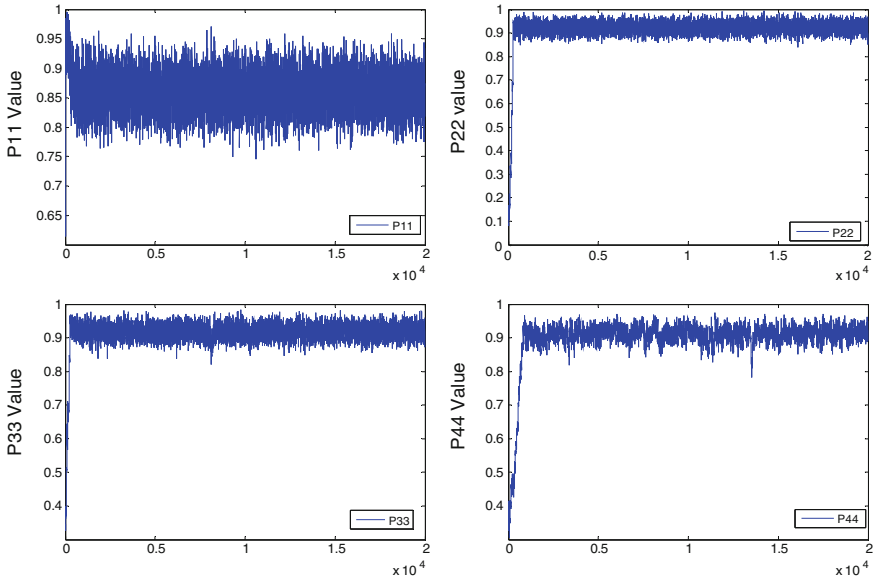


Fig. 1 Trace plots for main transition probability values after first 20,000 iterations

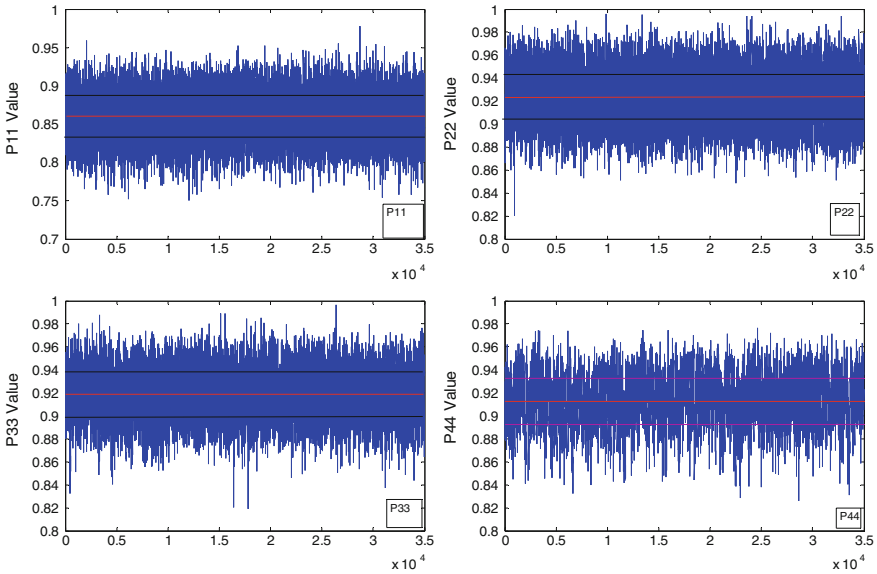


Fig. 2 Trace plots for up to 50,000 iterations after 15,000 warm-up runs for main TPM values

Table 1 Main transition probability values and 95 % confidence intervals for the entire dataset

p value	Mean	STD	Upper limit	Lower limit
P11	0.864	0.02564	0.867	0.860
P22	0.921	0.01651	0.923	0.918
P33	0.923	0.0178	0.925	0.920
P44	0.913	0.0195	0.915	0.910

condition state vector, Markov equation is applied to obtaining the future condition state proportions with time elapsing as given in Fig. 4.

$$\mathrm{P} = \begin{bmatrix} 0.86 & 0.14 & 0 & 0 & 0 \\ 0 & 0.92 & 0.08 & 0 & 0 \\ 0 & 0 & 0.92 & 0.08 & 0 \\ 0 & 0 & 0 & 0.91 & 0.09 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{1.8}$$

$$\mathrm{P} = \begin{bmatrix} 0.7496 & 0.2492 & 0.0112 & 0 & 0 \\ 0 & 0.8464 & 0.1472 & 0.0064 & 0 \\ 0 & 0 & 0.8464 & 0.1464 & 0.0072 \\ 0 & 0 & 0 & 0.84281 & 0.1719 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{1.9}$$

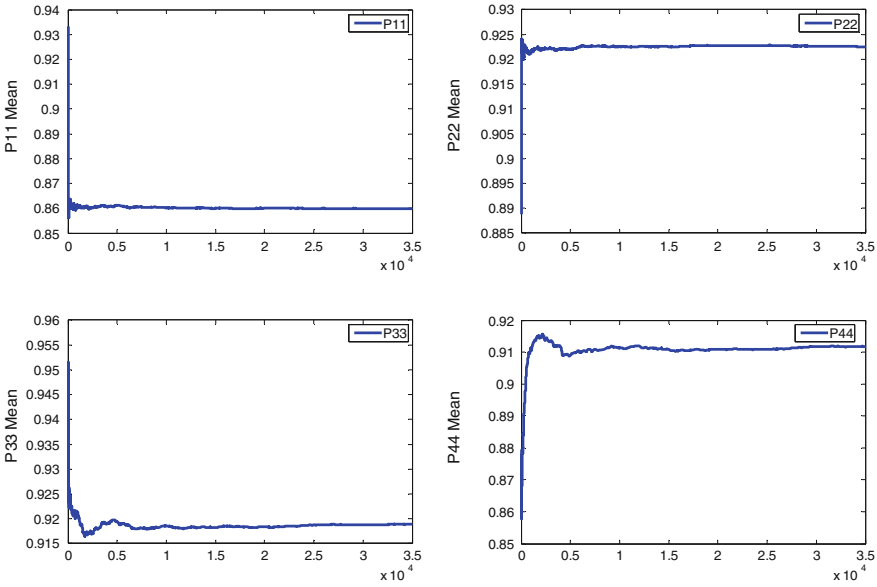


Fig. 3 Trace plots for mean transition probability values for up to 50,000 iterations after 15,000 warm-up runs

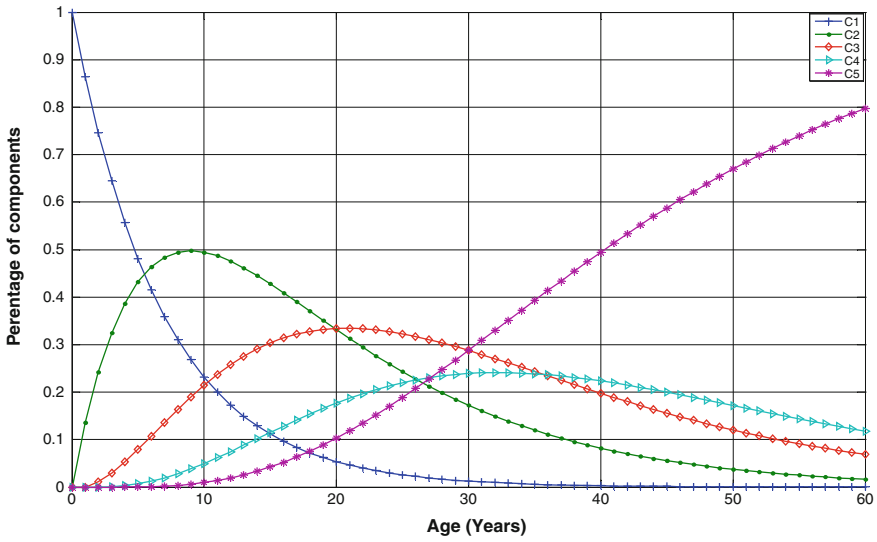


Fig. 4 Condition percentage distribution of each condition sate versus age in MCMC with MHA for the entire data set

Table 2 Chi-square values of MCMC method results

Method	Chi-Square values with 4 degrees of freedoms (≤9.49)	
	Calibrated data set	Test data set
MCMC simulation with MHA	0.268	1.508

4.2 Verification of Results

The Model was validated by using Goodness-of-fit test using Chi-squared test statistics (χ^2) which is based on null hypothesis that the observed number of bridge elements is matched with the predicted number of elements in different condition states [6]. The Chi-square values of the MCMC method for calibrated data set and test data set are 0.268 and 1.508, respectively; which are much smaller than the Chi-square critical value of 9.49, see Table 2. 95 % confidence level was used to evaluate the fitness of the models. Chi-squared test statistics (χ^2) for bridge element deterioration models in this study was calculated according to Eq. (1.10).

$$\chi^2 = \sum_{i=1}^{5} \frac{(O_i - P_i)^2}{P_i} \tag{1.10}$$

where, $\boldsymbol{O_i}$ is observed number of transoms in condition i (transom is the main element of transverse timber bridge deck), P_i is predicted number of transoms in condition i.

5 Conclusions

This chapter reviews the application of Markov Chain Monte Carlo (MCMC) approach with Metropolis Hasting Algorithm (MHA) for network level bridge deterioration modeling. From expert opinions and previous studies, contribution factors for rail bridge deterioration were identified. Bridge inventory data were collected from a main industrial partner in Australia and reviewed. Transition Probability Matrix (TPM) was estimated by using MCMC with MHA for bridge deck transoms with similar characteristics. The outcome of the MHA is sample distributions for transition probabilities which increases the chance of capturing true global optimum compared to Regression based NOA methods. The Output of the deterioration models were validated by using Goodness-of-fit test. According to Table 2, results show that Chi-square values of transition probabilities for calibrated and test data set are well below the limit values (Chi-square critical value). Obtaining very small Chi-square values compared to limit value convinced the superiority of the MHA and MCMC for bridge deterioration modeling. As given in

Table 1, the ability to express confidence intervals for transition probabilities is another advantage of MCMC method over conventional Markov calibration methods such as regression based and distribution based NOA. Major drawback of the proposed methodology over NOA is that MCMC seeks considerable number of condition rating data which expands the age range of the selected component group.

Further Work: Available Markov calibration techniques and MCMC will be applied to developing a network level deterioration models for other bridge components in order to make a further comparison with other methods.

Acknowledgments This research is funded by CRC for Rail Innovation Australia on Railway Asset Management Research Program.

References

1. Madanat S, Mishalani R, Wan Ibrahim W H (1995) Estimation of infrastructure transition probabilities from condition rating data. J Infrastruct Syst ASCE 1(2):120–125
2. Jiang Y, Saito M, Sinha KC (1988) Bridge performance prediction model using the Markov chain. Transportation Research Record 1180. Transportation Research Board, Washington, D.C., pp 25–32
3. Micevski T, Kuczera G, Coombes P (2002) Markov model for storm water pipe deterioration. J Infrastruct Syst ASCE 8(2):49–56
4. Nielsen D, Chattopadhyay G, Dhamodharan R (2012) Life cycle management of railway bridges: defect management. In: Conference on railway engineering, 10–12 Sept 2012, Brisbane, Australia, pp 425–434
5. Morcous G (2006) Performance prediction of bridge deck systems using Markov chains. J Perform Construct Facil 20:146–155
6. Tran HD (2007) Investigation of deterioration models for stormwater pipe systems. School of Architectural, Civil and Mechanical Engineering. Victoria University, Australia
7. Veshosky D, Beidleman CR (1996) Closure to "Comparative Analysis of Bridge Superstructure Deterioration". J Struct Eng 122(6):710–771

The Stress Dependence of the Magnetic Characteristics of Heat Resistant Steel 13CrMo4-5 and the Possibility of the Stresses Assessment on the Base of These Characteristics

D. Jackiewicz, J. Salach, R. Szewczyk and A. Bieńkowski

Abstract Paper presents the results of investigation on the tensile stresses dependence of magnetic characteristics of the heat resistant 13CrMo4-5 steel. For this investigation, the frame-shaped samples were used. Due to the specialized force reversing system, compressive force generates the uniform tensile stresses in the sample. Magnetic characteristics are measured under these stresses by digitally controlled hysteresis graph. On the base of results of measurements the magnetoelastic characteristics of resistant 13CrMo4-5 steel were determined. These characteristics indicate that change from elastic to plastic deformation significantly changes the magnetic properties of this steel. This information has great technical importance from the point of view of non-destructive testing of construction elements made of heat resistant 13CrMo4-5 steel.

Keywords Magnetoelastic effect · Heat resistant steel · Stress assessment

1 Introduction

Heat resistant 13CrMo4-5 steel is commonly used as a material for construction of critical elements of energetic infrastructure such as overheated steam pipelines [1]. Due to the fact, that malfunction of such pipelines may lead to serious conse-

D. Jackiewicz (✉) · J. Salach · R. Szewczyk · A. Bieńkowski
Institute of Metrology and Biomedical Engineering, Boboli 8, 02-525 Warsaw, Poland
e-mail: d.jackiewicz@mchtr.pw.edu.pl

J. Salach
e-mail: j.salach@mchtr.pw.edu.pl

R. Szewczyk
e-mail: szewczyk@mchtr.pw.edu.pl

A. Bieńkowski
e-mail: a.bienkowski@mchtr.pw.edu.pl

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_48

quences, both for the people's safety and economy, state of material of these construction have to be intensively monitored.

Among different available methods of non-destructive testing, magnetic properties oriented methods have significant advantages [2, 3]. First of all, non-destructive tests may be realized during the operation of pipelines, which reduces costs of maintenance. Moreover, magnetic tests (based on magnetoelastic characteristics of the material) are contact-less, which simplifies the process of tested element surface preparation [4, 5]. In addition, magnetic field generation, in the range of energy and frequency used for non-destructive tests, doesn't create health risk for operator, which is significant advantage in comparison with use of X-ray radiation.

However, magnetoelastic characteristics oriented methods of non-destructive tests are not commonly used in industry. The main barrier for such industrial application is the lack of knowledge about magnetoelastic characteristics [6] of specific types of steels used in energetic industry, such as heat resistant 13CrMo4-5 steel. This lack is directly connected with the lack of robust, unified methodology of testing the magnetoelastic characteristics of industrial types of steel.

This paper is trying to fill both of these gaps. It presents industrial application oriented methodology of magnetoelastic testing of frame-shaped samples made of different types of steels. Moreover, results of tests on heat resistant 13CrMo4-5 steel are also presented together with guidelines for stress assessment.

2 Method of Investigation

The frame-shaped sample used for magnetoelastic tests is presented in Fig. 1. On the columns of the sample both sensing and magnetizing windings were made. It is highly recommended to wound magnetizing and sensing windings on both

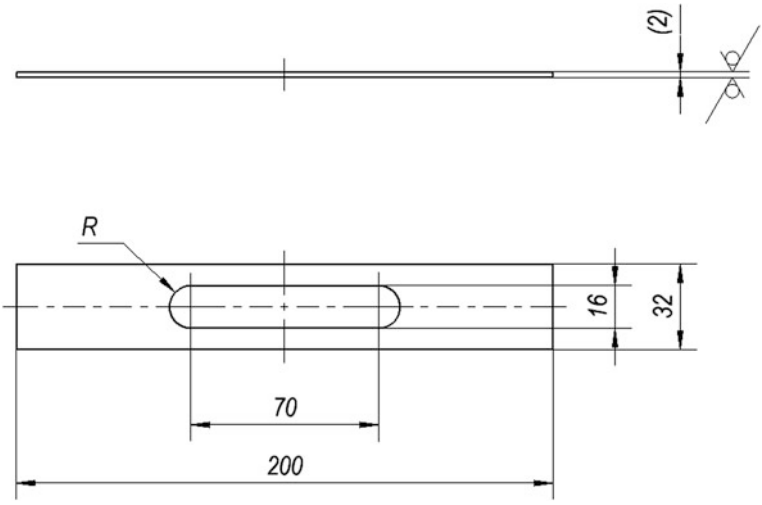


Fig. 1 Frame-shaped sample for the magnetoelastic tests

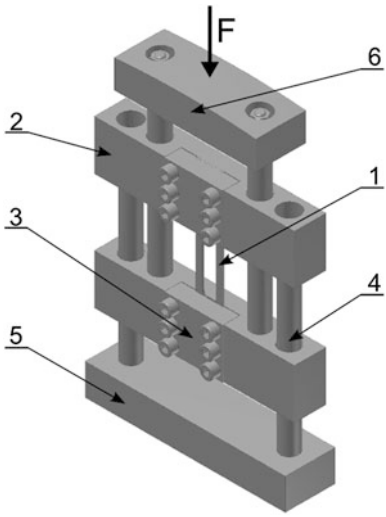


Fig. 2 Mechanical setup for testing the magnetic and magnetoelastic properties of frame-shaped samples: F—compressive force, *1*—tested frame-shaped sample, *2*—moving bar, *3*—sample holder, *4*—cylindrical columns, *5*—base of the device, *6*—upper bar

columns. Moreover, sensing winding should be located under the magnetizing winding to reduce demagnetization effects. In presented research, sample was wound by 700 turns of magnetizing winding (350 turns on each column) as well as 200 turns of sensing winding (100 turns on each column of the frame-shaped sample). Calculation of effective magnetic path length as well as effective magnetic cross-section of the frame-shaped sample was done according to "Calculation of the effective parameters of magnetic piece parts" [7].

Figure 2 presents the general view of mechanical setup for testing the magnetic and magnetoelastic properties of frame-shaped samples. With use of this system, the compressive force F can be converted to uniform tensile stresses in the columns of tested frame-shaped sample. It should be indicated, that precisely controlled compressive force F can be easily generated by e.g. oil press.

The schematic block diagram of computer controlled system for testing the magnetic and magnetoelastic properties of frame-shaped samples is presented in Fig. 3. The magnetizing current is generated by the KEPCO BOP36-6 voltage-current converter controlled by personal computer with National Instruments data

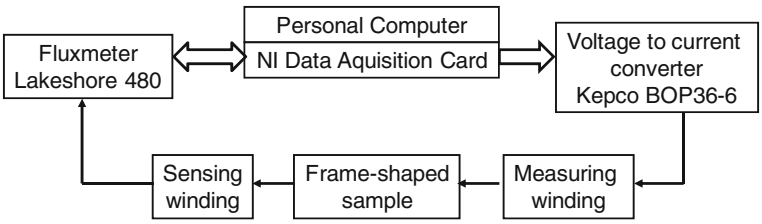


Fig. 3 Schematic block diagram computer controlled hysteresis graph system for magnetic and magnetoelastic testing

acquisition card. Current drives the magnetizing winding of frame-shaped sample, whereas signal from sensing winding is connected to the sensing input of the Lakeshore 480 fluxmeter. There is a two-way data transmission with fluxmeter: configuration is provided whereas measuring data set is acquired. Whole system is controlled by the hysteresis graph software developed in LabView environment.

3 Results

Figure 4 presents the experimental results of measurements of stress dependence of magnetic characteristics of frame-shaped samples made of heat resistant 13CrMo4-5 steel. Stress dependence of the shape of magnetic hysteresis $B(H)$ loops may be observed for different values of amplitude of magnetizing field H_m. It may be noted changes of the basic magnetic parameters: flux density, remanence, coercivity. From the point of view of utility and technical, changes of flux density and coercivity are the most interesting.

Figure 5 presents the magnetoelastic $B(\sigma)_{Hm}$ characteristics, whereas Fig. 6 presents the stress σ dependence of coercive field H_m. Under the tensile stresses value of flux density B in the sample first increase, and then, after reaching the Villari point it starts to decrease. It should be indicated, that this decrease starts to be the most significant for stresses σ about 300 MPa, which are connected with change from elastic to plastic deformation of sample made of heat resistant 13CrMo4-5 steel. Moreover, these changes are relatively higher for lower values of amplitude of magnetizing field H_m. This occurs due to the fact, that for lower values of magnetizing field H_m, participation of magnetoelastic energy in the total free energy is significantly higher.

Similar phenomena may be observed on stress σ dependences of coercive force H_c presented in Fig. 6. After reaching stresses σ connected with plastic deformation, value of coercive force H_c starts to change rapidly. This effect is connected with the hardening of the heat resistant 13CrMo4-5 steel under plastic deformation.

4 Possibility of the Stress Assesment

Presented experimental results indicate, that due to appearance of Villari point [8] on magnetoelastic characteristics of heat resistant 13CrMo4-5 steel, both the magnetoelastic $B(\sigma)_{Hm}$ characteristics as well as $H_c(\sigma)_{Hm}$ characteristics are not monotonous. For this reason, small values of tensile stresses σ can't be clearly assessed on the base of these characteristics.

However, for tensile stresses σ about 300 MPa, which are in the range near the change from elastic to plastic deformation, both flux density B and coercive force H_c start to change rapidly, giving clear and reliable signal of this change. This signal is very important from the point of view of non-destructive testing and stress assessment

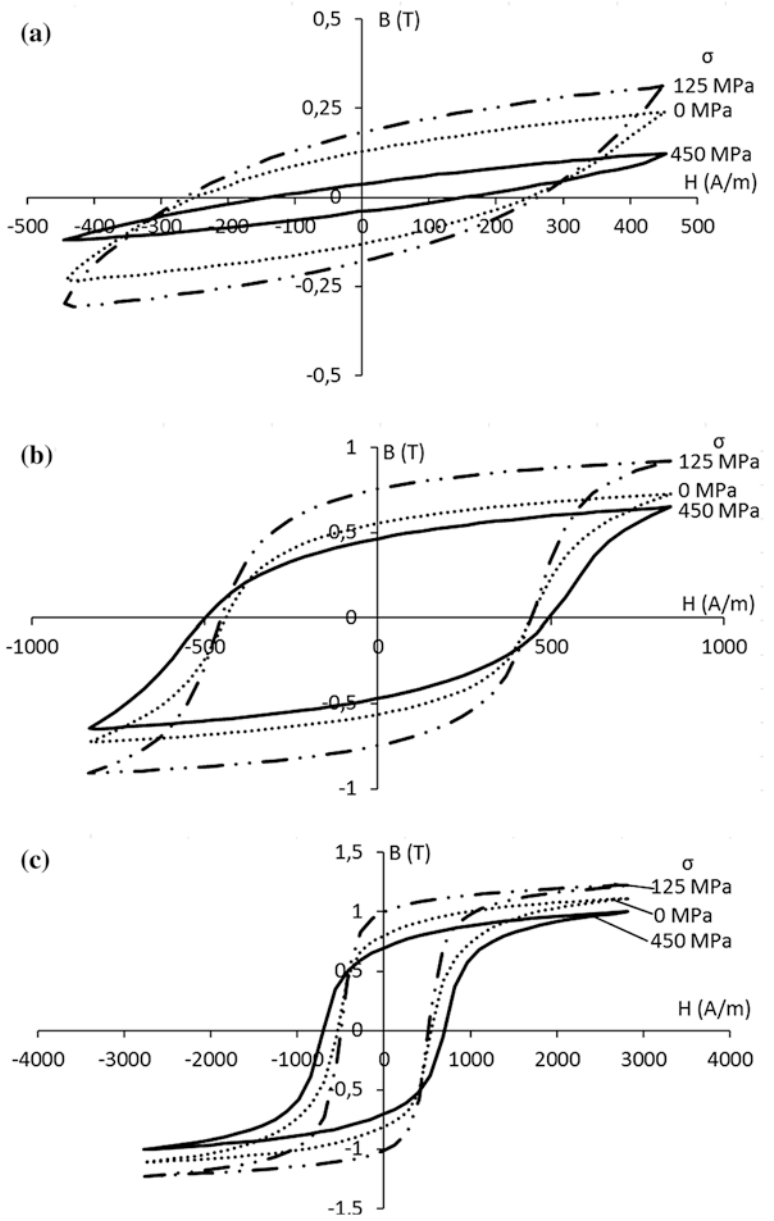


Fig. 4 The tensile stresses dependence of magnetic *B(H)* characteristics of frame-shaped samples made of heat resistant 13CrMo4-5 steel, for the three amplitudes H_m of magnetizing field: **a** H_m = 480 A/m, **b** H_m = 900 A/m, **c** H_m = 3,000 A/m

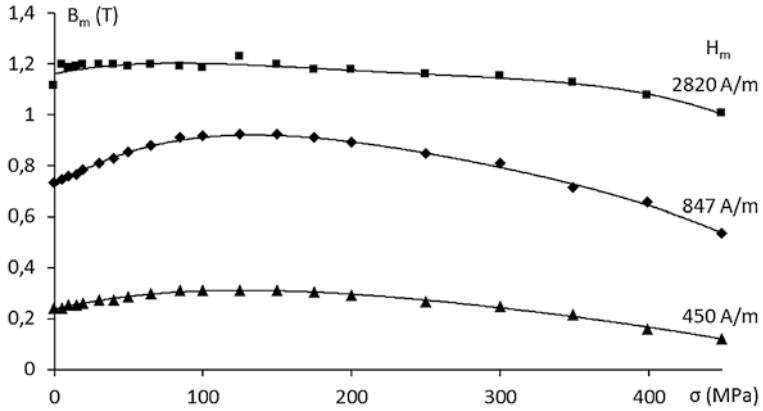


Fig. 5 The tensile stresses σ dependences of flux density B in heat resistant 13CrMo4-5 steel, for three value of amplitude of magnetizing field H_m

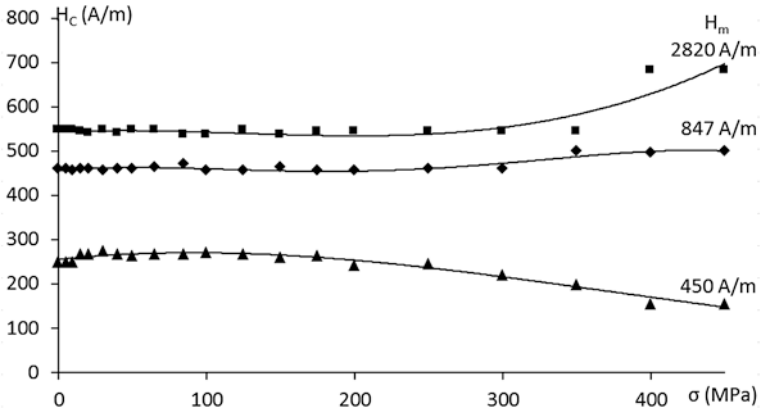


Fig. 6 The tensile stresses σ dependences of coercive force H_c in heat resistant 13CrMo4-5 steel, for three value of amplitude of magnetizing field H_m

in the construction elements made of 13CrMo4-5 steel. This phenomena is especially important in the case of contactless, non-destructive testing of construction elements of critical energetic infrastructure such as overheated steam pipelines.

5 Conclusion

Presented method of magnetoelastic testing of frame-shaped samples made of construction steels opens the new possibility of filling the gap connected with the lack of information about their magnetoelastic characteristics. With the use of this

method, the database covering wide variety of steels may be developed, creating the industry-applicable possibility of non-destructive tests of construction elements.

Presented results indicate, that magnetoelastic characteristics of heat resistant 13CrMo4-5 steel don't enable small values of tensile stresses σ assessment. However, for larger values of tensile stresses, which are in the range near the change from elastic to plastic deformation, both flux density B and coercive force H_c start to change rapidly, giving reliable signal, which is important from the point of view of non-destructive testing. For this reason, presented experimental results confirm feasibility of use of magnetoelastic effect in non-destructive testing of construction elements made of heat resistant 13CrMo4-5 steel.

Acknowledgments This work was partially supported by The National Centre of Research and Development (Poland) within grant no. PBS1/B4/6/2012.

References

1. Dobrzanski J, Sroka M (2008) Automatic classification of the 13CrMo4-5 steel worked in creep conditions. J Achiev Mater Manuf Eng 29(2):147–150
2. Xu B, Li HY (2012) Application of magnetoelastic effect of ferromagnetic material in stress measurement. Adv Mater Res 496:306–309
3. Xiao-yong Z, Xiao-hong Z (2012) Feature extraction and analysis of magnetic non-destructive testing for wire rope. In: Third International Conference on Digital Manufacturing and Automation, 2012, pp 418–421
4. Lei Ch, Xiangyu L, Tangsheng Y (2010) New magneto-elastic sensor signal test and application information computing and applications. Commun Comput Inf Sci 106:212–219
5. Wichmann HJ, & Holst A, Budelmann H (2009) Magnetoelastic stress measurement and material defect detection in prestressed tendons using coil sensors. In: NDTCE'09, Non-destructive testing in civil engineering, France
6. Szewczyk R, Svec P Sr, Svec P, Salach J, Jackiewicz D, Bienkowski A, Hosko J, Kaminski M, Winiarski W (2013) Thermal annealing of soft magnetic materials and measurements of its magnetoelastic properties. Meas Autom Robot 2:513–518
7. EN 60205:2006 Calculation of the effective parameters of magnetic piece parts
8. Szewczyk R, Bienkowski A, Kolano R (2003) Influence of nanocrystalization on magneto-elastic Villari effect in $Fe_{73.5}Nb_3Cu_1Si_{13.5}B_9$ alloy. Cryst Res Technol 38(3–5):320–324

Lithium-Ion Battery Degradation Related Parameter Estimation Using Electrochemistry-Based Dual Models

Yangbing Lou, Xiaoning Jin, Jun Ni, Sheng Cheng and X. Jin

Abstract This chapter presents an adaptive model for estimating the State of Charge (SOC) of a lithium-ion (Li-Ion) battery cell throughout its lifetime and its parameters based on electrochemical model. A Dual Extended Kalman Filter (DEKF) model is proposed for SOC estimation by using two cooperating extended Kalman filters, where the first one is responsible for estimating the SOC while the second one estimates the cell parameters indicating the level of cell deterioration due to aging. The dual filter combination is capable of tuning Kalman gains and providing accurate estimates even when the dynamics of the parameters change as the cell ages (e.g., inner resistance, capacity). By comparing with the experimental data, the results from the proposed method show an efficient SOC estimation with quick convergence and robust estimation of parameter changes in the long run.

List of Symbols

Symbol	Description (Unit)
a	Active surface area per electrode unit volume (cm^{-1})
c_i	Li-ion concentration (mol cm^{-3})
n	Particle coefficient (−)
r	Particle radius (cm)
A	Area (cm^2)
D	Diffusion coefficient (−)
E	Open circuit voltage (V)
F	Faraday's constant (C mol^{-1})
I	Current (A)
J	Butler–Volmer current density (A cm^{-3})

Y. Lou · X. Jin (✉) · J. Ni
University of Michigan, Ann Arbor, MI 48109, USA
e-mail: xnjin@umich.edu

S. Cheng
Beijing Shenzhou Aerospace Software Technology Company, Beijing, China

X. Jin
7, 14195 Berlin, Germany

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_49

K_{ef}	Effective electrolyte phase diffusion conductivity ($cm^2\ s^{-1}$)
L	Kalman Gain (−)
R	Gas constant ($J\ K^{-1}\ mol^{-1}$)
R_f	Film resistance (Ω)
T	Temperature (K)
α	Change transfers coefficients (−)
δ	Thickness (cm)
η	Overpotential (V)
θ	Normalized concentration (−)
ϕ	Potential (V)
$\sum$	Covariance (−)

1 Introduction

Lithium-ion (Li-ion) batteries are regarded as the most promising power energy storage technology for new generation electric vehicles. Compared to alternative battery technologies, Li-ion batteries have much higher energy density, exhibit no memory effect and greater durability. However, due to the complexity, safety concerns of Li-ion batteries and customer using habit, difficulty arises in properly estimating the battery energy level, e.g., state of charge (SOC), monitoring the cell degradation processes, and predicting their remaining useful life [1]. These features can be analyzed by model-based methods. The literature of Lithium ion battery modeling is generally categorized into two classes: (1) equivalent circuit models [2]; and (2) physics-based models built upon electrochemical reaction [3]. The equivalent circuit models have limited usefulness for large scale energy applications (e.g., electric vehicles) which require higher accuracy compared to portable electronic applications, especially during operations involving both micro-cycling and deep cycling. In addition, a large number of parameters are required to develop the equivalent circuit model and simulate the complete battery behavior [3]. Furthermore, since the physics parameters turn into mere fitting parameters for the equivalent circuit model, the intuition inside a battery is lost. Physics-based electrochemical models, including more detailed electrochemical phenomena in modeling, can not only resolve the above mentioned difficulties, but can also improve the estimation and prediction performance of battery cell. Many studies have developed simplified electrochemical-based models that can provide robust and efficient estimation of battery cell state and parameters without loss of computational efficiency [4, 5].

Precise estimation of SOC—a key battery indicator—is required by the electric vehicle application. SOC indicates how much power a battery has before it needs to be recharged [6]. SOC can reflect the energy level, performance, and determine other output, such as estimated voltage. In addition, the SOC of a battery cell needs

be determined to facilitate safety and efficiency during charging and discharging processes because accurate SOC estimation can help prevent over-charging and over-discharging conditions, and prevent different kinds of damage to the battery, and eventually extending the lifetime [7]. A number of techniques have previously been proposed to measure or estimate the SOC of battery single cell, each having its relative strength and limitation, as reviewed in [8, 9].

However, current existing literature on electrochemistry based SOC estimation focuses on single cell SOC estimation techniques for a short term, which could be inaccurate throughout the life time of the cell. Under different operating conditions, such as temperature, depth of charge and discharge, aging effects on each individual cell can be different in terms of inner resistance, capacity and dynamics of chemical reactions. Besides, the driving behavior may also cause different aging process among cells [10]. All these differences may gradually increase along the time and will be reflected by SOC divergence or the internal resistance divergence [11]. On the other hand, there are literatures showing the battery long-term degradation, such as equivalent circuit based estimation to determine charge capacity fade and internal resistance increment as a degradation indicator [12]; a charge-discharge capacity fade model based on the loss of active lithium ions due to solvent reduction reaction [13], but they were either based on simulation without real data verification, or lost most the actual information, which can be used to interpret the physical meaning of degradation process. There are few literatures showing the long-term cell degradation with electrochemistry based model and verified by experimental results, hence, there is no sufficient information for an estimation of available energy and power and the level of cell deterioration indicated by changes in physical parameters over time. Therefore, for a reliable and accurate battery management system, the changes of SOC and parameters of a battery cell need in the long term need to be accurately estimated for degradation monitoring. Methods have been provided, but with problems such as intensive computation, difficulty in online implementation in an automotive embedded system, and inaccuracy due to model constraints [14, 15].

This chapter proposes a method to estimate both the SOC and long-term cell parameters of the battery by integrating a simplified electrochemical battery model and a proposed Dual Extended Kalman Filter technique. The SOC and the film resistance of a single cell over its life time are estimated. The advantages of this proposed method are two-fold: (1) implementing physics-based models to provide physical interpretation of Lithium ion battery cell, and (2) utilizing dual models to maintain the long-term accuracy of estimates. The estimation result from this method can be further extended to battery performance prediction and health management by analyzing more long-term related parameters.

The remainder of the chapter is organized as follows. Section 2 describes the simplified electrochemical battery model which is built upon the LiCoO2 chemical reaction and lithium ion diffusion mechanism. Section 3 briefly reviews the Extended Kalman filter techniques and how it can be applied to the electrochemical model based SOC estimation. In Sect. 4, we propose a new dual-EKF model and demonstrate the parameters estimation over a long period of time as a second

Kalman filter in the dual model. The experimental data of battery test are compared with the results of the proposed model in Sect. 5. Section 6 concludes the chapter. The first lines of all subsequent paragraphs are ("Normal" style).

2 Electrochemical Battery Model

2.1 Chemical Reaction

The electrochemical battery model has been studied in the field of electrochemistry. In this chapter, the model is established based on the major components of positive material Cobalt: $LiCoO_2$. The overall chemical reaction for Li-ion battery formula is given by

$$LiCoO_2 + C \underset{\text{discharge}}{\overset{\text{charge}}{\rightleftharpoons}} Li_{1-k}CoO_2 + Li_kC \tag{1}$$

which can be derived into two electrode reactions. For positive side, Li^+ ions are extracted from $LiCoO_2$ by oxidation during charging and are inserted into $LiCoO_2$ by reduction during discharging.

$$LiCoO_2 \underset{\text{discharge}}{\overset{\text{charge}}{\rightleftharpoons}} Li_{1-k}CoO_2 + kLi^+ + ke^- \tag{2}$$

For negative side, Li^+ particles are inserted into Li_kC by reduction during charging and are extracted from Li_kC by oxidation during discharging.

$$kLi^+ + C + ke^- \underset{\text{discharge}}{\overset{\text{charge}}{\rightleftharpoons}} Li_kC \tag{3}$$

The *mol* fraction of Li^+ can be considered as the critical state of charge of the Li-ion battery, which is the electrode-average solid concentration at the electrolyte interface and can be transferred into the normally mentioned state of charge.

2.2 Li-Ion Diffusion and Concentration

The electrochemical principles are used to construct a physics-based model of a Li-ion battery. The one-dimensional model of a Li-ion battery considers the dynamics along only one axis (the horizontal *X*-axis) and neglects the dynamics along the rest two axes (*Y*-axis and *Z*-axis) [16]. This approximation is applicable to most cell structures with a large cross-sectional area and small currents.

For example, the characteristic length scale of a typical Li-ion cell along the X-axis is in the scale of 100 μm, whereas the characteristic length scale for the remaining two axes is on the order of 0.1 mm or more. To simplify the model, it is assumed that the average electrolyte concentration, c_e, is a constant. This assumption can be verified in [17] and is justified due to the insignificant difference (<5 %) observed in the electrolyte concentration in the battery [18].

The Li-ion diffusion in electrolyte and electrode can be descried by Fick's first and second law.

$$J_i(x,t) = -D_i \frac{\partial c_i(x,t)}{\partial x} \tag{4}$$

$$\frac{\partial c_i(x,t)}{\partial x} = D_i \frac{\partial^2 c_i(x,t)}{\partial x^2} \tag{5}$$

Since the concentration in electrolyte is assumed to be constant along the x direction, the electrode diffusion in one dimension is only considered. By considering the material diffusion inside representative solid material particles for each electrode, the system from $x = 0$ to $x = r$ is divided into spatial elements of the thickness Δr. J_i is the flux of the diffusing elements at location x and time t, which can be determined by current I and its corresponding locations. D_i is the diffusion coefficient and $c_i(x, t)$ is the concentration of the diffusing element i at location x and time t. Initial and boundary conditions have to be defined for each diffusion problem in (8) and (9). In a linear diffusion, we have discretized form

$$\partial N_i = AD(c_i - c_{i-1})/\partial x \tag{6}$$

where N is the number of moles within the element and A is the area. Based on the spherical geometry, we can obtain

$$\frac{\partial c_i}{\partial t} = \frac{D}{\Delta r^2}\left(\frac{i-1}{i} c_{i+1} - 2c_i + \frac{i+1}{i} c_{i-1}\right) \tag{7}$$

At the two boundaries, we have

$$\frac{\partial c_1}{\partial t} = \frac{D}{\Delta r^2}(2c_2 - 2c_1) \tag{8}$$

$$\frac{\partial c_{n-1}}{\partial t} = \frac{D}{\Delta r^2}\left(\frac{n-2}{n-1} c_{n-2} - \frac{n-2}{n-1} c_{n-1}\right) + \frac{DJ}{\Delta r}\frac{n}{n-1}. \tag{9}$$

Finally, state space equation for linear diffusion is

$$\dot{\boldsymbol{c}} = \frac{D}{\Delta r^2} A_s \boldsymbol{c} + \frac{D}{\Delta r}\frac{n}{n-1} B_s J \tag{10}$$

The parameter A_s and B_s are determined by Eq. (9). This approximation leads to an average value of the solid concentration that can be related with the definition of battery SOC. Although this simplified model results in loss of information, it can be efficient in control and estimation applications and still maintains a connection with the physical phenomena and dimensions.

The SOC is determined by the stoichiometry value θ_i in Eqs. (11) and (12) [19], where $\bar{c}_s$ is the average solid concentration and $c_{s\max}$ is the maximum solid concentration. Here, we assume that the maximum solid concentration is constant along the battery life cycle

$$SOC(t) = (\theta - \theta_0)/(\theta_1 - \theta_0) \tag{11}$$

$$\theta = \frac{\overline{c_s}}{c_{s\max}} \tag{12}$$

2.3 Butler-Volmer Current, Overpotential and Voltage Computation

The overall battery terminal voltage V is constructed in Eq. (13) by battery's open circuit voltage (OCV, E_{ocv}), overpotential (η), electrostatic potentials (ϕ), and film resistance (R_f) on the electrodes surface. The details can be found in [3, 18]. We also applied the average model in developing the electrochemical mechanism.

$$V = E_p - E_n = (E_{ocv,p} + \eta_p + \phi_p) - (E_{ocv,n} + \eta_n + \phi_n) - IR_f \tag{13}$$

The OCV can be determined based on empirical correlation described in [20]. The overpotential can be determined by Butler-Volmer current density while the electrostatic potentials are determined by the thickness of electrodes and separator. By substituting Eqs. (14–18) into (13), we can develop the battery voltage.

$$\phi_p - \phi_n = \frac{\delta_p + 2\delta_{sep} + \delta_n}{2Ak_{ef}} \tag{14}$$

$$E_{ocv,p} = v_0 + v_1\theta + v_2\theta^{0.5} + v_3\theta^{-1} + v_4\theta^{1.5} + v_5 e^{v_6\theta + v_7} + v_8 e^{v_9\theta + v_{10}} \tag{15}$$

$$E_{ocv,n} = u_0 + u_1\theta + u_2\theta^2 + u_3\theta^3 + u_4\theta^4 + u_5\theta^5 + u_6\theta^6 + u_7 e^{u_8\theta} \tag{16}$$

$$\eta_p - \eta_n = \frac{RT}{\alpha nF} \ln\left(\frac{j_p + \sqrt{j_p^2 + 4a^2 j_0^2}}{j_n + \sqrt{j_n^2 + 4a^2 j_0^2}}\right) \tag{17}$$

$$j_0 = (c_e(c_{s\max} - c_{se})c_{se})^{\alpha} \tag{18}$$

J_0 is the referenced current density determined by Li^+ concentrations at different stage [18]. a is the Active surface area per electrode unit volume, and v and u is the coefficients for empirical correlation. Besides, δ is the thickness of electrode and separator. R is the inner resistance, R_f is the film resistance, A is the electrode plate area, and k_{eff} is the effective electrolyte phase diffusion conductivity [21].

Thus, the average Butler–Volmer current considers a representative solid material particle somewhere along the negative (n) and positive electrode (p). This simplified model has similarities with the "single-particle" model introduced in [16]. The diffusion dynamics are approximated with a state space equation of first order ordinary differential equation, which has been described in above section, Eq. (10). Furthermore, in the electrode-average model, the cell voltage depends, through E_p and E_n, on the solid-electrode concentration instead of the average single-particle bulk concentration.

3 Extended Kalman Filter

Extended Kalman filtering (EKF) is widely used to estimate system state and parameters for non-linear cell models by using a linearization process at every time step. The EKF method is able to automatically compute the dynamic battery cell "state" and its error bounds in real time based on real real-time measurements. Usually, battery voltage, current, and temperature are measured with sensor noise. In this study, instead of deriving the SOC as the system state in EKF model directly, we use the solid concentration at the electrodes (c_{es}) as the system state to be determined. Solid concentration c_{es} is determined by the approximation analysis of Li-ion diffusion process in the previous section. The electrochemical model involves both system noise and sensor noise, w and v, which are assumed to be zero-mean, Gaussian noises, respectively.

In Sect. 2, the voltage is found to be a nonlinear function of Li-ion concentration. We use the voltage as the output in the EKF model. The EKF can then be implemented as:

$$\dot{x} = Ax + Bu + L(y - \hat{y}) \tag{19}$$

$$\hat{y} = V(x, u) \tag{20}$$

where $V(x, u)$ is determined by the solid concentration x (c_{se}) of the final segment in Eq. (13). A and B are the same matrix in Eq. (10). The procedure for Kalman gain calculation is shown below [22].

Initial condition (k = 0):

$$x_0^+ = x_{initial}$$

$$\Sigma_{x,0}^+ = (x_0 - x_0^+)(x_0 - x_0^+)^T$$

For k = 0, 1, 2, 3 ...

Time update:

State: $x_{k+1}^- = Ax_k^+ + Bu_k$

Error covariance: $\Sigma_{x,k+1}^- = A_k \Sigma_{x,k}^+ A_{k-1}{}^T + \Sigma_w$

Kalman: $L_{k+1} = \Sigma_{x,k+1}^- C_{k+1}^T [C_{k+1} \Sigma_{\hat{x},k+1}^- C_{k+1}^T + \Sigma_v]^{-1}$

Measurement update:

State: $x_{k+1}^+ = Ax_{k+1}^- + L_{k+1}[y_{k+1} - (Ax_{k+1}^- + Bu_k)]$

Error covariance: $\Sigma_{x,k+1}^+ = (1 - L_{k+1} C_{k+1}) \Sigma_{x,k+1}^-$

Here, $\sum$w and $\sum$v are the covariance of noise w from system, and noise v from sensor. And the matrix C_k for the nonlinear system is shown below. The full derivation is not presented due to the equation complexity.

$$C_k = \frac{\partial V}{\partial x_k}|_{x_k = x_k^-} \tag{21}$$

$$C_k = \frac{\partial V}{\partial c_{s,p,(N-1)}} = \frac{\partial E_p}{\partial c_{s,p,(N-1)}} - \frac{\partial E_n}{\partial c_{se,n}} \frac{\partial c_{se,n}}{\partial c_{s,n,(N-1)}} \tag{22}$$

Figure 1 illustrates the discrete time EKF model. x and $\hat{y}$ are the estimated state and output, respectively. The input u is the current density in this case. For software computation implementation, the system is further transformed into discrete time form, which is shown below.

4 Dual Extended Kalman Filter (DEKF)

The state of charge (SOC), as one of the descriptive quantities of the present system state changes rapidly, while others may change very slowly with time, such as inner resistance and cell capacity, which might change as little as 30 % during 1,000

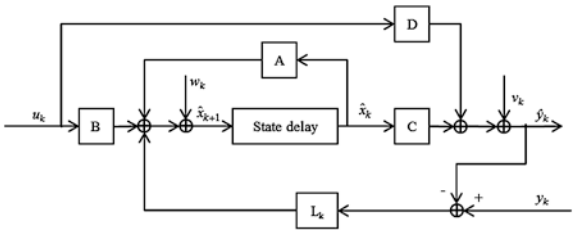


Fig. 1 Discrete time system with EKF

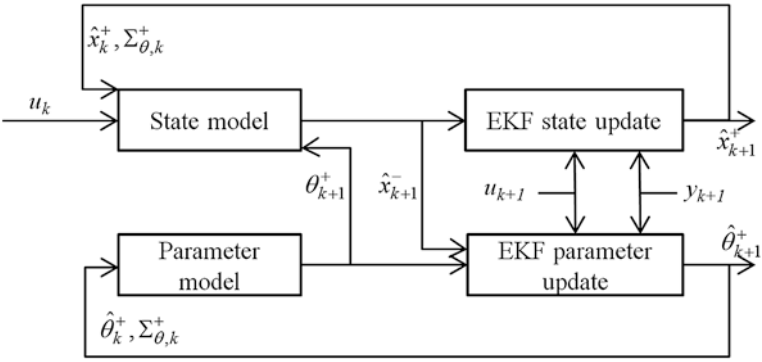


Fig. 2 Discrete time system with DEKF

cycles [13]. These parameters decaying over time are often used to describe the state-of-health, and are important for cell degradation analysis and remaining useful life estimation.

The DEKF method is employed here to help estimate the values of state and parameters simultaneously, where two Kalman filters are implemented in the system shown below. The structure of the dual model is also described in Fig. 2.

$$y_{k+1} = g(x_k, u_k, \theta_k) + v_k \tag{23}$$

$$z_{k+1} = g(x_k, u_k, \theta_k) + e_k \tag{24}$$

In this model, a critical cell parameter, cell film resistance, is tracked. The first system is described in Sect. 3, the second system in the discrete time form is determined below. Equation (23) is the same equation as Eq. (13), based on the concentration.

$$\theta_k = R_{f,k} \tag{25}$$

$$R_{f,k+1} = R_{f,k} + r_k \tag{26}$$

$$z_{k+1} = g(x_k, u_k, R_k) + e_k \tag{27}$$

Equations (25) and (26) show that the film resistance is generally time invariant, but it may vary slowly during long term due to gradual cell degradation during cycling. The process is modeled by r_k as the small degradation step. The output equation for the state-space model of true parameter dynamics is the cell output voltage estimate plus the estimation error e_k. Hence, in the dual model, voltage is the measurement for output update.

5 Experiments and Discussion

In this section, we present the experiments of the $LiCoO_2$ battery cell test and estimation results based on the proposed DEKF model. A single cell was tested in a climate chamber for 450 cycles. The battery temperature was maintained between 20 and 29 °C, while most of the time, the temperature is kept at 22 °C. Due to the small variant, the temperature change has insignificant effect on the battery performance. The battery nominal voltage is 3.6 V and for each discharging cycle, the battery was discharged from 4.1 V (terminal voltage) to 2.5 V (terminal voltage), with a constant 1.5 A current. For each charging cycle, the constant current constant voltage (CCCV) method was applied to protect battery from high voltage and overcharging damage. During the cycle, the battery was charged back to 4.2 V (terminal voltage) by using constant 1.5 A current and then was kept at constant voltage by reducing the current. In order to see the degradation effects over the multiple cycles, we plot the battery measurement in the 1st cycle and the 401st cycle, as well as the SOC estimation in Figs. 3 and 4. During the experiment, the current of discharging process or charging process were both recorded as positive values. High accuracy measurement sensors were applied to record the voltage and current.

From Fig. 3, it is shown that at 401st cycle, the discharging time was reduced by 21 %. This discharging capacity fading phenomenon can be viewed as one of the main indication of battery aging effect. Furthermore, the nominal voltage for the 401st cycle was also smaller than the one for the 1st cycle. This phenomenon can be physically described by film resistance increment on the surface of electrode particles. The long-term capacity fade and film resistance increment will be further discussed later.

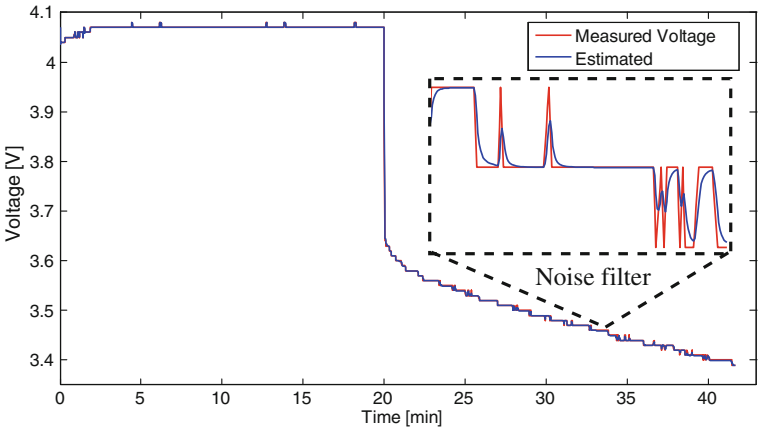


Fig. 3 Short term voltage estimation

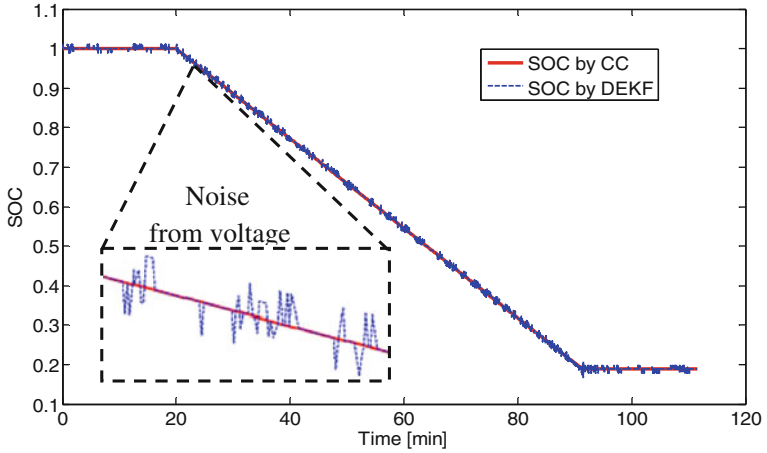


Fig. 4 Short term SOC estimation

The result of short-term voltage and SOC estimation in one discharging cycle is shown in Figs. 4 and 5. In Fig. 4, the estimated voltage by using DEKF is compared with the measurement. It can be observed that the estimate converges to the measurement and the noise variant is reduced by implementing Kalman Filter. In Fig. 5, the SOC estimation is further compared with *coulomb counting* (CC) method. It can be observed that the estimated SOC converges to the CC results. During our experiment, the current was precisely controlled at 1.5 A and the SOC determined by CC method is a linear line with no noise. We noted that the noise of SOC estimated by DKEF is introduced by measured voltage. Here, we use the CC method conducted in ideal experimental environment to verify the DEKF method. It can be inferred that when there is noise in the current measurement in real applications, the DKEF SOC estimation will be more effective than CC SOC estimation due to its noise filtering capability.

The discharging capacity for each cycle determined based on CC methods and DEKF SOC changes are shown in Fig. 6. Though some large battery capacity recoveries are observed due to experiment interruption and variants, it can be concluded that the discharging capacity determined based on DEKF is valid and effectively monitor the degradation process.

In order to understand the cell degradation effects on the film resistance, we use DEKF to estimate the long-term film resistance change. Since there is no way to directly measure the film resistance, an alternative method, *film growth rate method* (FGRM) [^{i}] based on a first-principal battery model is used to compare and verify the estimation performance of the proposed DEKF method. FGRM has been considered as the most efficient method in determining the film growth at solid electrolyte interphase (SEI) which is one of the main contributors to capacity fade and battery age [20]. The single cell film resistance growth is determined by

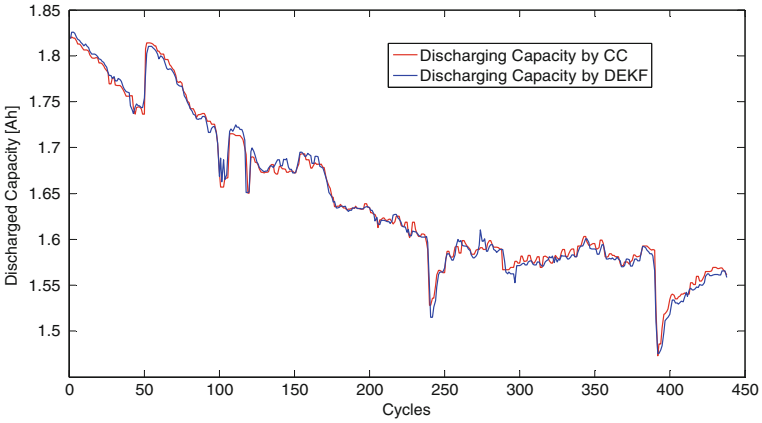


Fig. 5 Long-term capacity fade estimation

$$R_{film}(x,t) = R_{SEI} + \delta_{film}(x,t)/\kappa \tag{28}$$

$$\frac{\partial \delta_{film}(x,t)}{\partial t} = -\alpha J(x,t) \tag{29}$$

where R_{SEI}, κ and α are the battery parameters. δ is the thickness of the layer. The comparison of per unit area film resistance determined by DEKF and the FGRM is shown in Fig. 6. The film resistance is averaged in each cycle to better demonstrate our result. In Fig. 6, we observe that the film resistance growth estimates have similar values and same pattern as the FGRM. However, there is limitation by

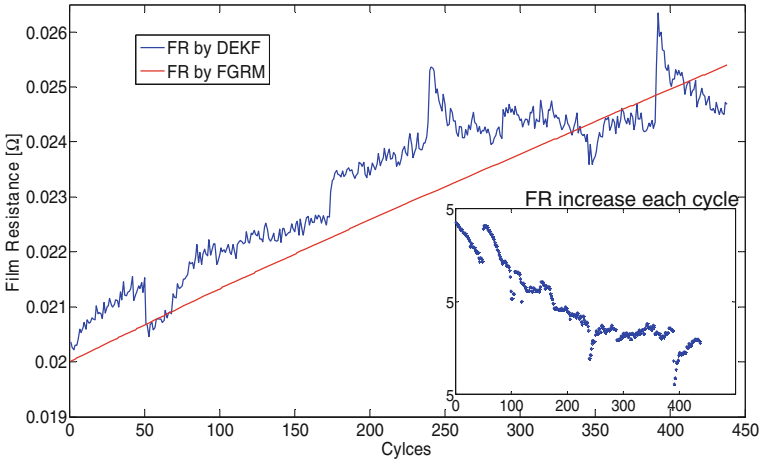


Fig. 6 Long-term film resistance estimation

applying FGRM to verify our simplified battery model due to some assumptions in the simplified electrochemical model. It is assumed that the maximum solid concentration is constant over the battery life cycle and the SOC is determined based on the 1st cycle of the battery. Due to the degradation effects, the speed of voltage drop during discharging will actually increase along its life cycle. The voltage drop effects involve several parameter changes including film resistance, diffusion coefficient, material thickness and lithium concentration changes, but only the film resistance is considered in our simplified electrochemical model. Therefore, the film growth rate estimated upon the assumption might overestimate the film resistance. It is also shown that the estimation by DKEF has larger values than FGRM. One possible reason is that the first dynamic model embedded in the DKEF involves many parameters other than film resistance. Some of the parameters will also change though the cell life time, instead of keeping constant in the present model. This simplification could affect the accuracy of estimated parameter.

One advantage of the proposed DEKF over the FGRM is that DEKF can also be applied to estimate other cell parameters, even if we don't know the parameter aging mechanism, such as lithium ions concentration reduction [14]. In other words, this comparison envisions that the DKEF is a useful technique for parameters tracking in the long run throughout the cell aging process, especially when the fundamental aging mechanisms of the parameters are not clearly understood.

6 Conclusions and Future Work

This chapter investigates a method for estimating both the state of charge and cell parameters of a Li-ion battery cell over its lifetime using Dual Extended Kalman Filters. To determine the state of charge accurately, the electrochemical model is developed to represent the cell dynamics, which is more advanced than equivalent circuit model. The experimental results and simulation results show an efficient estimation and quick convergence of the SOC and robust estimation of parameters changing in the long run. Experiment results and estimation results show that the discharging capacity and film resistance determined based on DEKF is valid and effectively represent the degradation process. It can also be inferred that if there is noise in the sensor measurement and battery system itself, the estimations based on DEKF will be more accurate than that based on CC method due to the noise filtering capability and long-term parameter tuning function.

Future work might investigate more cell parameters other than the film growth rate, provided that we have a better understanding of complicated aging effects on different parameters. A future study might also consider abrupt changes of parameters for accurate battery performance prognostics. The accuracy of the battery state and parameters estimation by applying DKEF can be further improved at the cost of increased model complexity and computational effort.

7 Acknowledgments

This research is funded by NSF Industry and University Cooperative Research Program (I/UCRC) for Intelligent Maintenance Systems (IMS) at the University of Michigan. The authors would like to thank IMS center at the University of Cincinnati for providing experimental data.

References

1. Armand M, Tarascon J (2008) Building better batteries. Nature 451(7179):652–657
2. He H, Xiong R, Guo H, Li S (2012) Comparison study on the battery models used for the energy management of batteries in electric vehicles. Energy Convers Manag 64:113–121
3. Chaturvedi NA, Klein R, Christensen J, Ahmed J, Kojic A (2010) Algorithms for advanced battery-management systems, IEEE Control Syst Mag 30(3):49–68
4. Sikha G, White R, Popov B (2005) A mathematical model for a lithium-ion battery/electrochemical capacitor hybrid system. J Electrochem Soc 152(8):A1682–A1693
5. Smith KA, Rahn CD, Wang C (2010) Model-based electrochemical estimation and constraint management for pulse operation of lithium ion batteries. IEEE Trans Control Syst Technol 18 (3)
6. He W, Williard N, Chen C, Pecht M (2012) State of charge estimation for electric vehicle batteries under an adaptive filtering framework. In: Proceedings of IEEE 2012 prognostics and system health management conference, PHM-2012
7. Cheng K, Divakar BP, Wu H, Ding K, Ho H (2011) Battery-management system (BMS) and SOC development for electrical vehicles. IEEE Trans Veh Technol 60(1):76–88
8. Zhang J, Lee J (2011) A review on prognostics and health monitoring of Li-ion battery. J Power Sources 196(15):6007–6014
9. Samadani S, Fraser R, Fowler M (2012) A review study of methods for lithium-ion battery health monitoring and remaining life estimation in hybrid electric vehicles, SAE technical paper, 2012-01-0125. doi:10.4271/2012-01-0125
10. Sarre G, Blanchard P, Broussely M (2004) Aging of lithium-ion batteries. J Power Sources 127(1–2):65–71
11. Dubarry M, Vuillaume N, Liaw B (2010) Origins and accommodation of cell variations in Li-ion battery pack modeling. Int J Energy Res 34(2):216–231
12. Plett GL (2004) Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs. Part 3. State and parameter estimation. J Power Sources 134(2):277–292
13. Ning G, Popov BN (2004) Cycle life modeling of lithium-ion batteries. J Electrochem Soc 151 (10):A1584–A1591
14. Hu C, Youn B, Chung J (2012) A multiscale framework with extended Kalman filter for lithium-ion battery SOC and capacity estimation. Appl Energy 92:694–704
15. Farag M, Ahmed R, Gadsden SA, Habibi SR, Tjong J (2012) A comparative study of Li-ion battery models and nonlinear dual estimation strategies. In: 2012 IEEE Transportation electrification conference and expo (ITEC), p 8
16. Danilov D, Niessen RAH, Notten PHL (2011) Modeling all-solid-state Li-ion batteries. J Electrochem Soc 158(3):A215–A222
17. Santhanagopalan S, White RE (2006) Online estimation of the state of charge of a lithium ion cell. J Power Sources 161:1346–1355
18. Vidts PD, Delgado J, White RE (1995) Mathematical modeling for the discharge of a metal hydride electrode. J Electrochem Soc 142(12):4006–4013

19. Domenico DD, Stefanopoulou A, Fiengo G (2010) Lithium-ion battery state of charge and critical surface charge estimation using an electrochemical model-based extended Kalman filter. J Dyn Syst Meas Control 132(6):061302
20. Doyle M, Fuentes Y (2003) Computer simulations of a lithium-ion polymer battery and implications for higher capacity next-generation battery designs. J Electrochem Soc 150: A706–A713
21. Domenico DD, Fiengo G, Stefanopoulou A (2008) Lithium-ion Battery state of charge estimation with a Kalman filter based on a electrochemical model. In: Proceedings of the IEEE international conference on control applications, pp 702–707
22. Plett GL (2004) Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs. Part 1. Background. J Power Sources 134(2):252–261
23. Moura SJ, Forman JC, Bashash S, Stein JL, Fathy HK (2011) Optimal control of film growth in lithium-ion battery packs via relay switches. IEEE Trans Ind Electron 58(8):3555–3566

Segregation of Close Frequency Components Based on Reassigned Wavelet Analysis for Machinery Fault Diagnosis

Ahmed M. Abdelrhman, M. Salman Leong, Lim Meng Hee and Salah M. Ali Al-Obaidi

Abstract Vibration signals of rotating machinery often contain many closely located frequency components. While Fast Fourier Transform (FFT) analysis of the signals can identify exact frequency components in the vibration spectrum easily, conventional wavelet analysis is generally incapable of discriminating closely located frequency components in vibration signals due to overlapping and interference appearing in wavelet results. Wavelet transforms based on wavelet reassignment algorithm to improve time-frequency resolution display is presented in this chapter. The proposed reassigned (modified) Morlet wavelet was tested using simulated signal and experimental data obtained from a multi-stage blades rotor test rig. This study showed that this method was capable of segregating close BPF components which were otherwise lumped together in conventional wavelet analysis display. The reassigned Morlet wavelet analysis was shown to be useful for multi stage blade rubbing diagnosis as well as other general condition monitoring applications such as those for gear and bearing faults diagnosis.

Keywords Frequency components separation · Morlet wavelet · Machinery faults diagnosis

A.M. Abdelrhman (✉) · M.S. Leong · S.M. Ali Al-Obaidi
Institute of Noise and Vibration, UTM, Jalan Semarak, 54100 Kuala Lumpur, Malaysia
e-mail: ahmedrabak@yahoo.com

M.S. Leong
e-mail: salman@ic.utm.my

S.M. Ali Al-Obaidi
e-mail: salahmtc@yahoo.com

L.M. Hee
UTM Razak School of Engineering and Advanced Technology, Jalan Semarak,
54100 Kuala Lumpur, Malaysia
e-mail: mhlim@ic.utm.my

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_50

1 Introduction

Vibration analysis is widely used in machinery diagnosis; of which the most commonly used signal processing method involves transformation of the vibration signal from time domain to frequency domain to determine fault signatures in the FFT spectrum. The vibration signals often contain many closely located frequency components which can be easily seen in the conventional vibration FFT spectrum. A short coming with examination of the vibration spectrum is the loss of time information of the signal. Wavelet analysis technique in this case provides an alternative in the representation of the signal's contents with time-frequency display. Wavelet analysis however has its own pitfall as it is difficult to analyse signals with close frequency components which would appear completely overlapped in a wavelet map. Close frequency overlapping caused by frequency interference and spectral smearing in the results representation can lead to interpretation difficulties [1–3]. Minimizing overlapping and spectral smearing is therefore very important for clear and precise signal representation; as well as increasing the time and frequency resolution for enhancing fault detection capability. Wavelet reassignment technique has been widely used to improve the wavelet analysis resolution in time frequency plane. Peng et al. [4] for example used reassigned wavelet scalograms for better modal parameter estimation; while Sun et al. [5] used the reassigned wavelet scalogram for close vibration mode identification in structural system. In this chapter reassigned Morlet wavelet analysis is proposed to segregate close frequency components of vibration signals with an intent to achieve better concentration of signals energy distribution in both time and frequency planes.

2 Wavelet Transform

One of the shortcomings of Fourier transform technique is that the analysis of signals with a constant resolution and a single window used for all frequencies. Wavelet transforms was introduced to overcome these limitations by using multi-resolution approach with different window functions (son wavelets) for each associated frequency in the signal. Wavelets in general can be categorized into two types: discrete wavelet transforms (DWT) and continuous wavelet transforms (CWT). For machinery vibration type fault diagnosis, CWT is often used [6]. As stated above, wavelet transforms differ from the FFT technique as it analyze the signals in a more flexible way with a variable window width. This feature allows wavelet to have high signals localization in the time-frequency plane. The flexibility of the wavelet function allows it to change the mother wavelet shape in the time frequency plane becoming tall and thin at high frequencies and short and wide at lower frequencies. Another important distinction between wavelet and FFT analysis is that wavelet analysis is not limited to the use of sinusoidal analyzing functions only, but also employ a large selection of localized waveforms or wave function as long as they satisfy the predefined mathematical criteria [7].

2.1 Continuous Wavelet Transform

The continuous wavelet transform for any signal $x(t)$ is given by Eq. (1)

$$W(a,b) = <\psi_{a,b}(t), x(t) \geq = |a|^{-1/2} \int x(t)\psi^*_{a,b}dt \quad (1)$$

where, $W(a, b)$ is the wavelet transform, a is the scale parameter and b is the time location parameter. Both parameters may vary continuously, $\psi_{a,b}$ mother wavelet, factor $a^{-1/2}$ is used to ensure energy preservation during transformation.

The wavelet transform $W(a, b)$ is a function which provides detailed information for the signal $x(t)$ at different levels of resolution by shifting the parameter b for each scale a. It can also determine how much the signal $x(t)$ is similar to the daughter wavelet function $\psi_{a,b}(t)$ at different scales [8]. CWT has two magnificent advantages: the absence of any artificial cross-components, and is easy and simple to be adapted to any time and frequency resolution [9]. On the other hand, the basic design of continuous wavelet normally produces fine frequency resolution and coarse time resolution at low frequencies; and fine time resolution with coarse frequency resolution at high frequencies. It is deemed to be impossible to achieve both high resolutions in time and in frequency at the same window defined in the time-frequency plane [10]. It is always a balance between both time and frequency resolutions. The generated result of some inspected signals such as those contains close frequency components are therefore unreadable and difficult to be interpreted correctly as the results is not detailed enough, and suffer from lack of sufficient resolutions as close frequency components used to appear in the wavelet map overlapped and interfere with each other.

2.2 Morlet Wavelet

There are many types of mother wavelet available that could be selected. The selection of a proper and suitable wavelet function for specific signal thus is very important as it is the key factor for successful and accurate results. The Morlet wavelet function is frequently used [10–15] in machinery condition monitoring and fault diagnosis, typically to extract out faults features. The literature has reported Morlet having better similarity to vibration signals in comparison to many other functions such as Daubechies (1–43), Coiflet, Symlet, Gaussian, complex Gaussian and Meyer [16]. For many mechanical systems, impulses in the vibration signals are often symptoms of faults occurrence. Morlet wavelet is usually used to analyse this type of signals because it is very similar to this impulse components. Morlet wavelet is defined as in Eq. (2) as:

$$\Psi(t) = \exp\left(-\frac{\beta^2 t^2}{2}\right)\cos(\pi t) \tag{2}$$

The bandwidth parameter β of the Morlet wavelet can be adjusted to adapt to those impulses with any decaying rate. In addition the time-frequency resolution of Morlet wavelet can also be adapted to different signals of interest [11]. By dilation with scale a and translation with b, a son wavelet can be acquired as in Eq. (3) as:

$$\psi_{a,b}(t) = \exp\left[-\frac{\beta^2(t-b)^2}{a^2}\right]\cos\left[-\frac{\pi(t-b)}{a}\right] \tag{3}$$

It could be seen from Eqs. (2) and (3) that the Morelet wavelet is a cosine signal decaying exponentially on both sides having an impulsive shape. Figure 1 shows the Morelet mother wavelet.

These three parameters a, b and β are controls the time-scale plane of the computed wavelet. Different values for a, b and β correspond to different segmentations. The scale a estimated using Matlab™ function *SCAL2FRQ,* The syntax for this function is as in Eq. (4).

$$F = \text{scal2frq}\ (\text{A},'\,\text{wname}', \text{DELTA}) \tag{4}$$

where F is the desired frequency, A the calculated scale equivalent to the desired frequency, '*wname*' the wavelet name (*Morl*) and *DELTA* the sampling rate (*1/ sampling frequency*). The translation parameter b represents the data points or the inspected signal length. Parameter β controls the shape of the mother wavelet and balances the time resolution and the frequency resolution of the Morlet wavelet.

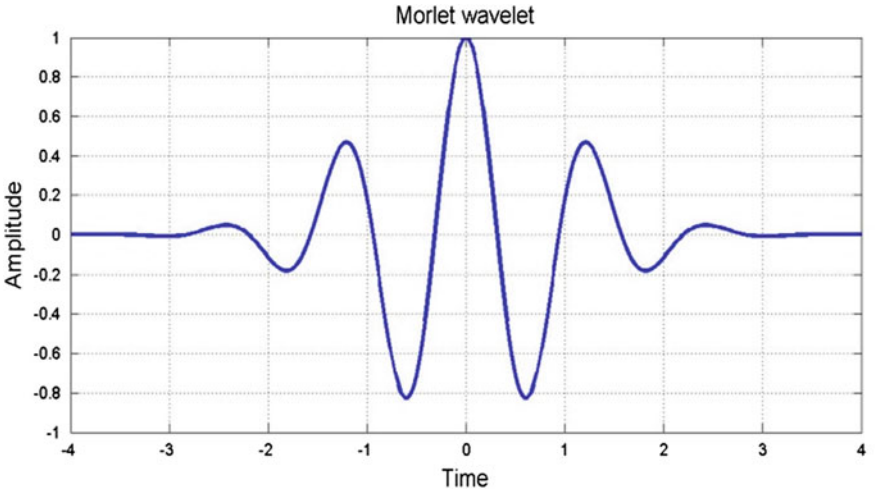


Fig. 1 Morelet mother wavelet

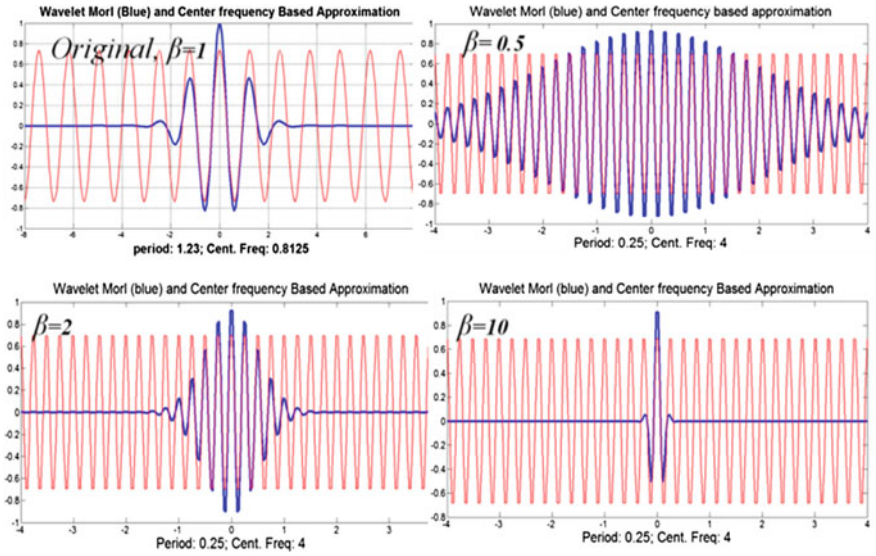


Fig. 2 Morlet wavelet shapes for different centre frequency and β values

When β decreased, the frequency resolution will increase accordingly. On the other hand, the time resolution will decrease. The finest frequency resolution of Morlet wavelet could be obtained when the bandwidth β tends to zero; as the Morlet wavelet becomes a cosine function. The finest time resolution of Morlet wavelet could be obtained when the bandwidth β tends to infinity, as Morlet wavelet becomes a *Dirac* function (8, 11). For each certain signal there is always an optimal value for the bandwidth β that exists for a best time frequency resolution.

Figure 2 shows Morlet wavelet shapes at two different centre frequencies (0.8125 and 4) and four different values of β (1,0.5, 2 and 10). These two parameters, centre of frequency and band-width efficiently control the shape of the mother wavelet (dilation or compression) and the time-frequency resolution of the Morlet wavelet. The Morlet wavelet can therefore be adapted to any signal by adjusting these two parameters [17]. The Morlet wavelet with central frequency less than 5 is more oscillatory, and allows for a better resolution of frequency components [7]. In this work, β = 0.5 and centre of frequency 4 was found to be the optimal values.

3 Signal Simulation Study

Signal simulation study was undertaken to investigate the effects of the reassignment method on results of vibration analysis. Signals consisting of three frequencies components (80, 110, and 150 Hz) with additive noise were generated using

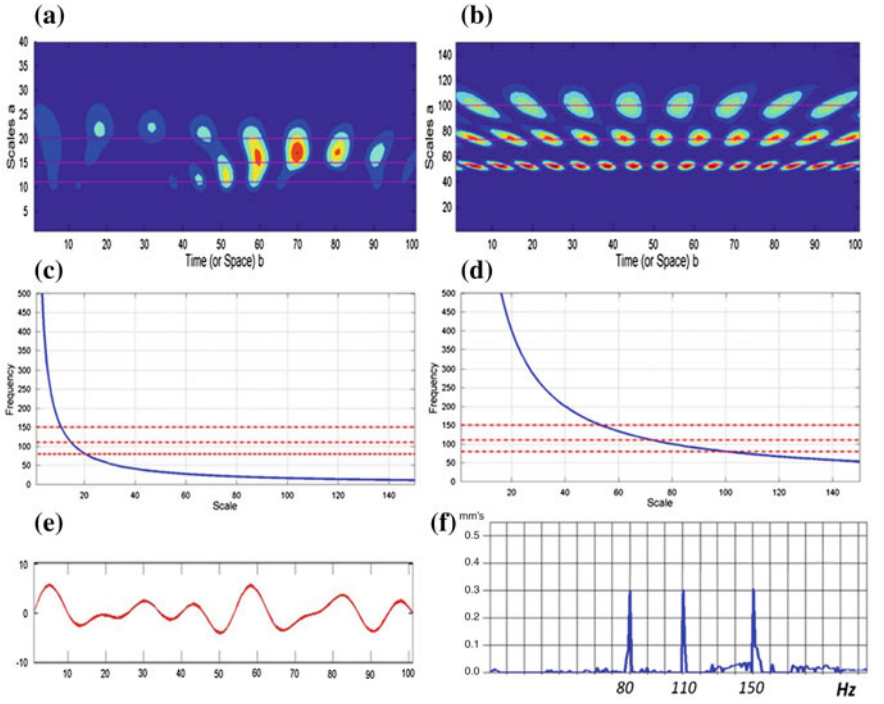


Fig. 3 Analysis results of the simulated signal. **a** Percentage of energy for each wayelet coefficient, **b** Percentage of energy for each wayelet coefficient, **c** Correlation of scales and frequencies, **d** Correlation of scales and frequencies, **e** Simulated signal, **f** Frequency spectrum

MATLAB at sampling frequency of 1,024 and time duration of 0.05 s. The results of simulated signal analyzed using Morlet wavelet and the reassigned Morlet wavelet are shown in Fig. 3a, b. Figure 3c, d show the correlation between scales and frequencies of these plots; while Fig. 3e shows the simulated time wave signal and Fig. 3f show the corresponding FFT plot.

From results obtained from the simulated signal, Fig. 3a, b gives comparison between the wavelet map of the reassigned Morlet wavelet (at centre of frequency = 4 and β = 0.5) and the original Morlet wavelet (at centre of frequency = 0.8125 and β = 1). The effects of the overlapping and interference could be seen in the plot of original Morlet wavelet in Fig. 3a. It can be seen that these three frequencies were indistinguishable and totally lumped together giving ambiguous analysis results with no any further details of the signal frequency components. On the contrary, the proposed method effectively separates the three frequency components of the signal in time frequency plane as evident in Fig. 3b. The reassignment technique dilates the mother wavelet to become more oscillatory providing better frequency resolution. On the other hand, the frequency components of 80, 110 and 150 Hz were located at very low scales of 20, 15 and 11 respectively

when using the original Morlet wavelet as seen in Fig. 3c; while it is located at higher scales of 100, 73 and 53 when using the proposed reassigned method, as seen in Fig. 3c.

4 Experimental Study with Multi-stage Rotor

In a study for blade rubbing diagnosis in rotating machinery a multi stage rotor system was fabricated. This test facility was used to examine the effectiveness of the proposed method for multi-stage blade rubs. A general view of the experimental test rig is shown in Fig. 4. The test rig consisted of three rows of rotor blades (with 8, 11 and 13 numbers of blades); with three additional rows of stator blade (with 12, 14 and 16 numbers of blades). The rotor is driven by an electrical motor via a variable speed controller.

Vibration signals were measured using tri-axial accelerometers mounted on the outer casing of the rotor system with sampling rates of 2 kHz using a multi-channel data acquisition system (IMC cs-3008). A tachometer was used for speed detection and once per revolution trigger.

Vibration signal measured from the rotor test rig at running speed of 1,200 RPM (20 Hz) were analyzed using both Morlet wavelet and the reassigned Morlet wavelet. Figure 5a, b illustrate the Morlet wavelet maps and reassigned Morlet wavelet maps of the signal in one cycle of rotation. Figure 5c, d shows the correlation between scales and frequencies. Figure 5e, f shows the time domain signal

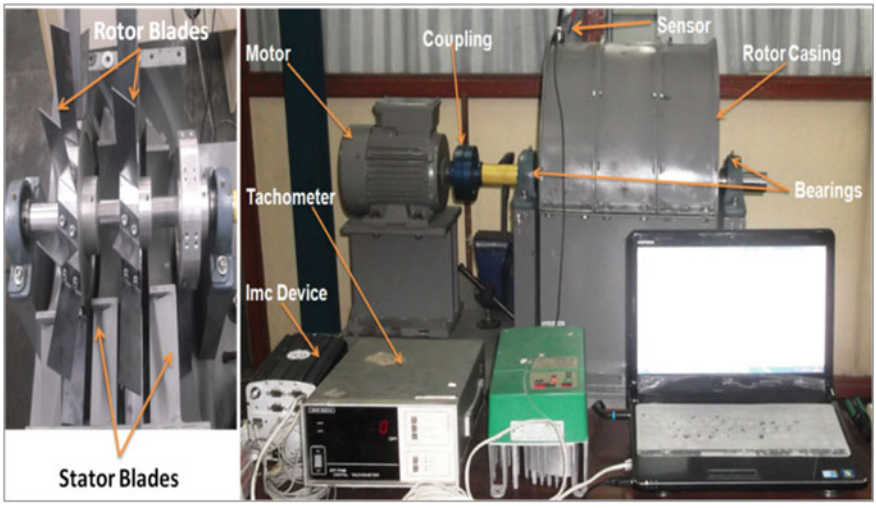


Fig. 4 Experimental test rig setup

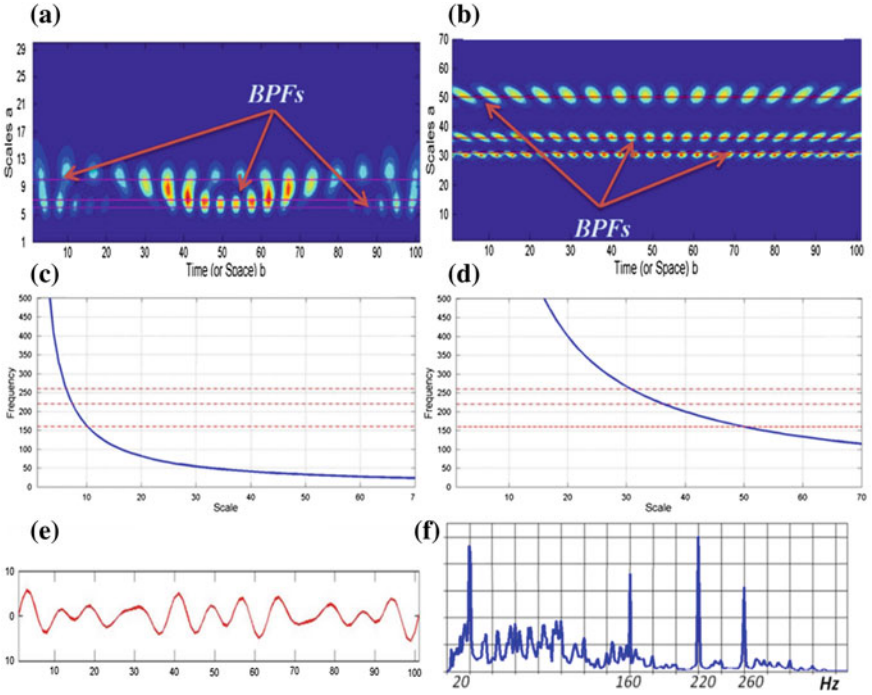


Fig. 5 Analysis results of the experimental signal. **a** Percentage of energy for each wavelet coefficient, **b** Percentage of energy for each wavelet coefficient, **c** Correlation of scales and frequencies, **d** Correlation of scales and frequencies, **e** Experimental signal, **f** Frequency spectrum

and its corresponding frequency spectrum of the acquired signal. The running frequency 20 Hz and the blade passing frequencies of the three stages 160, 220 and 260 Hz it can be seen clearly in the vibration spectrum.

It can also be seen from Fig. 5a, b that the proposed reassignment method was effective in segmenting the closely located frequency components and thus the individual blade passing frequencies of the rotor system could be seen easily as evident in Fig. 5a. In addition to this it was clear in Fig. 5c that the inspected signal and the BPF components were reconstructed in wider scale range as the three blade passing frequencies located at scales of 50, 37 and 31. It was located at 10, 7 and 6 scale values when using original Morlet wavelet.

Although results of Fig. 5 illustrate only the healthy rotor system, it is believed that any distortion in BPF due to blade fault could be seen from the reassigned wavelet map. Further work and experimental studies need to be conducted to confirm the above.

5 Conclusion

Wavelet analysis is becoming a widely used tool for machinery faults diagnosis. Due to the effects of frequency overlapping and interference, conventional wavelet analysis is however incapable of discriminating closely located frequency components. A method for closed vibration frequency components separation was proposed based on Morlet wavelet reassignments. This proposed technique provided better results for frequency segmentation of close frequency components when compared with conventional wavelet analysis. The centre of frequency and parameter β were found to be effective in adapting Morlet wavelet signals with increased wavelet resolutions. With better frequency resolutions it was thus possible to segregate frequency components of BPF for blade faults diagnosis in a multi-stage blades rotor.

References

1. Tse PW, Yang W-x, Tam H (2004) Machine fault diagnosis through an effective exact wavelet analysis. J Sound Vib277(4):1005–1024
2. Peng Z, Tse PW, Chu F (2005) An improved Hilbert-Huang transform and its application in vibration signal analysis. J Sound Vib 286(1):187–205
3. Abdelrhman AM, Leong MS, Hee LM, Ngui WK (2013) Application of wavelet analysis in blade faults diagnosis for multi-stages rotor system. Appl Mech Mater 393:64–959
4. Peng Z, Meng G, Chu F (2011) Improved wavelet reassigned scalograms and application for modal parameter estimation. Shock Vib 18(1):299–316
5. Sun Z, Hou W, Sun L (eds) (2006) Close-mode identification based on wavelet scalogram reassignment. In: 24th conference and exposition on structural dynamics 2006 (IMAC—XXIV); 2006 30 Jan–2 Feb 2006, St Louis, Missouri, USA
6. Tse P, Yang W (eds) (2002) A new wavelet transform for eliminating problems usually occurring in conventional wavelet transforms used for fault diagnosis. In: The Ninth international congress on sound and vibration, ICSV'9
7. Addison P, Watson J (2002) FENG T. Low-oscillation complex wavelets. J Sound Vib 254 (4):62–733
8. Zhang D, Sui W (eds) (2009) The optimal morlet wavelet and its application on mechanical fault detection. In: IEEE 5th international conference on wireless communications, networking and mobile computing, 2009 WiCom'09 2009
9. Liu J, Wang W, Golnaraghi F (2008) An extended wavelet spectrum for bearing fault diagnostics. IEEE Trans on Instrum Meas 57(12):12–2801
10. Peng Z, Chu F, Tse PW (2005) Detection of the rubbing-caused impacts for rotor–stator fault diagnosis using reassigned scalogram. Mech Syst Signal Process 19(2):391–409
11. Lin J, Qu L (2000) Feature Extraction based on Morlet wavelet and its application for mechanical fault diagnosis. J Sound Vib 234(1):48–135
12. Peng Z, Chu F, He Y (2002) Vibration signal analysis and feature extraction based on reassigned wavelet scalogram. J Sound Vib 253(5):100–1087
13. Shibin W, ZK Z, Yingping H, Weiguo H (2010) Adaptive parameter identification based on morlet wavelet and application in gearbox fault feature detection. EURASIP J Adv Signal Process
14. Junsheng C, Dejie Y, Yu Y (2007) Application of an impulse response wavelet to fault diagnosis of rolling bearings. Mech Syst Signal Process 21(2):9–920

15. Lin J, Zuo M (2003) Gearbox fault diagnosis using adaptive wavelet filter. Mech Syst Signal Process 17(6):69–1259
16. Rafiee J, Rafiee M, Tse P (2010) Application of mother wavelet functions for automatic gear and bearing fault diagnosis. Expert Syst Appl 37(6):79–4568
17. Jiang Y, Tang B, Qin Y, Liu W (2011) Feature extraction method of wind turbine based on adaptive Morlet wavelet and SVD. Renew Energy 36(8):53–2146

Feature Extraction of Rubbing Fault Based on AE Techniques

Wenxiu Lu and Fulei Chu

Abstract The rotor-to-stator rub is one of the main serious malfunctions that often occur in rotating machinery. The acoustic emission (AE) signal is very sensitive to rubbing occurrence and development. However, it is still very difficult to identify the rubbing AE signal from other faults AE signals such as crack AE signal, corrosion AE signal and so on. Rubbing AE signal is nonlinear and non-stationary, and Hilbert–Huang transform is powerful in processing nonlinear and non-stationary signals. Then Hilbert–Huang transform is used to extract the instantaneous frequency of the AE signal of rubbing fault. The experiment results show that the Hilbert–Huang transform has a good potential for the acoustic emission signal processing in rubbing fault diagnostics.

Keywords Rubbing fault · Acoustic emission signal · Hilbert–Huang transform · Instantaneous frequency

1 Introduction

Rotor-to-stator rub is a serious malfunction in rotating machinery. It often causes catastrophic failure and subsequent economic loss. It is very important and necessary to early detect and diagnose the rub-impact of rotating machinery timely and accurately, avoiding severe damage and expensive repairs.

A comprehensive research has been performed on the vibration of a rubbing rotor system. Muszynska [1] made a comprehensive review on this problem and gave a list of previous papers on the rub-related vibration phenomena during rubbing. Chu [2, 3] discussed the bifurcation and chaotic motion of a rub-impact

W. Lu (✉) · F. Chu
Department of Mechanical Engineering, Tsinghua Universtiy, Beijing 100084, China
e-mail: luwenxiu@mail.tsinghua.edu.cn

F. Chu
e-mail: chufl@mail.tsinghua.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_51

rotor system, and designed a rub-impact test stand and observes very rich forms of periodic and chaotic vibrations through experimental verification. However, when the rub is slight, the vibration behaviour of rotor system is almost the same to the normal rotor, that is, the vibration is not sensitive to slight rub. Acoustic emission technology is suitable technology to diagnose the slight fault. Acoustic emissions are defined as transient elastic waves generated from a rapid release of strain energy caused by a deformation or damage within or on the surface of a material [4]. Recent years, application of Acoustic emission technique has been growing in fault diagnostics of rotating machinery, especially in bearing defect diagnosis [5], gear fault detection [6] and rubbing fault diagnosis [7–9].

Sato [7] reported that AE measurement can provide a valuable complementary tool for diagnosing rubbing in fast rotating plant such as turbine generators. Wang and Chu [8] extracted the rub fault feature through AE signal waveform analysis technology, and developed the AE rubbing positioning method based on wavelet decomposition. Hall and Mba [9] proposed that modelling the cumulative distribution function of rub-induced AE signals with respect to appropriate theoretical distributions, and quantifying the goodness of fit with the Kolmogorov–Smirnov (KS) statistic, offers a suitable signal feature for diagnosis. These researches indicated that the AE technique may be used to detect rotor-stator rubbing and find the rub location. However, the time-frequency of AE wave is very complex. AE wave is a non-stationary stochastic signal, and the traditional data processing techniques such as Fourier Transform, Short-time Fourier Transform, Wavelet Transform and so on, may not be suitable for such non-stationary signal. Hilbert–Huang Transform (HHT) [10] is empirical, direct and adaptive, which is particularly suitable for non-stationary signal.

In this paper, a special stator is designed to simulate the rubbing fault, and the AE signal was recorded with the development of rubbing. Then HHT is used to extract the instantaneous frequency of the AE signal of rubbing fault. The experiment results show that the Hilbert–Huang transform has a good potential for the acoustic emission signal processing in rubbing fault diagnostics.

2 AE Analysis Technique Based on HHT

The Hilbert–Huang Transform was developed by Huang et al. in 1998 and the essence of HHT is to identify the intrinsic oscillation modes of time series through empirical mode decomposition (EMD). Generally, the components of shortest period at each instant will be identified and decomposed into the first Intrinsic Mode Function (IMF). And the components of longer periods will be identified and decomposed into the following IMFs in sequence. The IMFs have both variable amplitude and frequency as functions of time.

The IMF is defined as the following:

(1) In the whole data set, the number of extrema and the number of zero crossing must either equal or differ at most by one.

(2) At any point, the mean value of the envelope defined by local maxima and the envelope defined by local minimal is zero.

The instantaneous frequency and amplitude of IMF can be easily derived by Hilbert transform, and thus the Hilbert spectrum of IMF can be obtained.

The main idea of EMD is to determine the instantaneous equilibrium position by averaging the upper and lower extrema envelope, thereby to extract the IMF. The main steps can be the following:

(1) Identify the entire local extrema of original signal $x(t)$, and then connect all the local maxima by a cubic spline line as the upper envelope. Repeat the procedure for the local minimal to produce the lower envelope. The upper envelope and lower envelopes should cover all the data between them.
(2) The mean of the upper envelope and lower envelope is designed as $m(t)$
(3) The new time series $h(t)$ can be difference between the original signal and the mean value

$$h(t) = x(t) - m(t) \tag{1}$$

(4) If $h(t)$ is an IMF, then $h(t)$ is the first IMF of signal $x(t)$. If $h(t)$ is not an IMF, $h(t)$ is treated as the original signal $x(t)$ and repeat (1), (2) and (3) until $h(t)$ reaches an IMF. Therefore, the first IMF can be designated as:

$$c_1(t) = h(t) \tag{2}$$

$c_1(t)$ should contain the high frequency component of original signal.
(5) Separate $c_1(t)$ from $x(t)$

$$r_1(t) = x(t) - c_1(t) \tag{3}$$

(6) Treat $r_1(t)$ as the original signal and repeat above processes. The second IMF component $c_2(t)$ of $x(t)$ can be obtained. Repeat above processes n times, then the n-IMFs of signal $x(t)$ can be obtained:

$$\begin{aligned} r_2(t) &= r_1(t) - c_2(t) \\ r_3(t) &= r_2(t) - c_3(t) \\ &\cdots \\ r_n(t) &= r_{n-1}(t) - c_n(t) \end{aligned} \tag{4}$$

The decomposition process can be stopped when $r_n(t)$ becomes a monotonic function from which no more IMFs can be extracted. By summing up the extracted IMFs and $r_n(t)$, the original signal can be expressed as:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{5}$$

Thus, the original signal $x(t)$ is decomposed into n-empirical modes and a residue $r_n(t)$ which is the mean trend of $x(t)$.

The IMFs $c_1(t), c_2(t), \ldots, c_n(t)$ include different frequency bands ranging from high to low. The frequency components contained in each frequency band are also time-varying, which are very suitable for Hilbert transform. For one IMF $c_i(t)$, its Hilbert transform is as:

$$\widehat{c}_i(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{c_i(\tau)}{t - \tau} d\tau \tag{6}$$

With this definition, an analytic signal is expressed as:

$$z_i(t) = c_i(t) + i\widehat{c}_i(t) = a_i(t) e^{i\Phi_i(t)} \tag{7}$$

In which

$$a_i(t) = \sqrt{c_i^2(t) + \widehat{c}_i^2(t)} \tag{8}$$

$$\Phi_i(t) = \arctan \frac{\widehat{c}_i(t)}{c_i(t)} \tag{9}$$

The instantaneous frequency can be obtained as following:

$$\omega_i(t) = \frac{d\Phi_i(t)}{dt} \tag{10}$$

After performing the Hilbert transform to each IMF component, the original signal can be expressed in the following form:

$$x(t) = \sum_{i=1}^{n} a_i(t) e^{i\Phi_i(t)} = \sum_{i=1}^{n} a_i(t) e^{i \int \omega_i(t) dt} \tag{11}$$

Equation (11) enables us to represent the amplitude and the instantaneous frequency as functions of time in a three-dimensional plot, in which the amplitude can be contoured on the time-frequency plane. The time-frequency distribution of the amplitude is designated as the Hilbert spectrum $H(\omega, t)$.

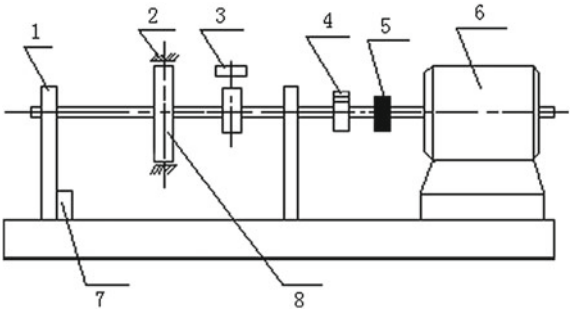


Fig. 1 Rotor rig. *1*. Bearing house, *2*. Stator, *3*. Eddy current transducer, *4*. Key-phasor, *5*. Shaft coupling, *6*. Motor *7*. AE sensor *8*. Rotor

1. Bearing house 2. Stator 3.Eddy current transducer
4. Key-phasor 5.Shaft coupling 6.Motor 7. AE sensor 8. Rotor

3 Experimental Set-up

The experimental rig is driven by a direct current shunt motor as shown in Fig. 1. The rotor speed can be adjusted from 0 to 10,000 rpm rapidly by the voltage speed controller. The experimental setup consists of the rotor kit with two supports and some disks, and a set of data acquisition system including AE sensors and key-phase sensor. The two bearings are oil film bearings and the length of the bearing house in axial direction is 25 mm. The diameter of the shaft is 9.5 mm with the length 500 mm. Any position along the shaft can be selected as bearing point.

The key-phasor is a closing sleeve with an axial slot and the key-phase signal is obtained by eddy current transducer. The AE sensors used for this experiment were broadband type sensors with a relative flat response in the region between 20 and 200 kHz (Physical Acoustic Corporation). The AE sensors are usually placed on the non-rotating member of machine, such as the bearing house or the pedest. All the signals from sensors were pre-amplified, filtered, and connected directly to a commercial data acquisition card, where a sampling rate of 1 MHz per channel was used during the test. The output signal from the AE sensors was pre-amplified at 40 dB.

4 Experimental Results and Discussion

A special structure of stator is designed as shown in Fig. 2 to simulation the rubbing fault. It is easy for the stator to be installed or detached. The clearance between the rotor and stator is adjustable to meet different experiment conditions. The rotor was started up from still to a certain speed, at which the rubbing had not occurred; then the speed was increased until the rubbing occurred; Keeping increasing the speed up to its first natural frequency, the rubbing was becoming severe, and the obvious impact voice could be heard. The severe rubbing state was kept for about 10 s, and then the speed was decreased to non-rubbing. The sample data was analyzed with MATLAB.

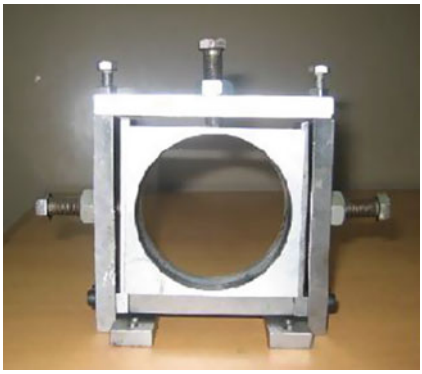


Fig. 2 Stator structure

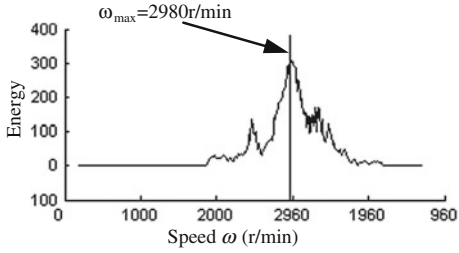


Fig. 3 Trend wave of AE energy

The energy of AE signal above the threshold (the threshold is set as 0.04 V) are observed as the rotating speed changes, as shown in Fig. 3. It can be clearly seen that the rub-impact occurs at ω = 1,870 r/min according to the AE signal. The rub-impact becomes severe with the increasing speed. At ω = 2,980 r/min, the rub-impact becomes the most severe. Subsequently, rub-impact becomes slight with the decreasing speed. Last, no rub-impact occurs when the rotor is away from the stator.

The waveform and spectrum of pulse AE signal of rubbing fault is shown in Fig. 4. HHT is applied to the pulse AE signal, and the signal is decomposed into 5 IMF components c_1–c_5, as shown in Fig. 4, which include different components from lower to high frequency. We can observe that the frequency of peak value in the spectra of c_4 and c_5 IMF is less than 20 k, which means the IMF component is distorted because the AE signal is filtered with a bandpass filter with passband 20–200 k. Then c_1, c_2 and c_3 IMF components are further analyzed through Hilbert–Huang spectrum, as shown in Fig. 5. At about 0.5 ms, the frequency from 0 to 200 kHz is visible from Fig. 5, which means the rub occurs at this instant. From the Hilbert–Huang spectrum of IMF c_3, it can be also seen that the AE events (rub) occur at time 1.5 and 1.9 m. From 0.6 to 1.5 m, the AE wave is propagating between the shaft, bearings and other components of rotor system. From the spectrum of IMF c_3, the propagation characteristics of rubbing AE signal can be more clearly observed. As indicated by the arrows, the interval time is about 0.1 m, which means the propagating time in the shaft is about 0.1 m. Then we can compute

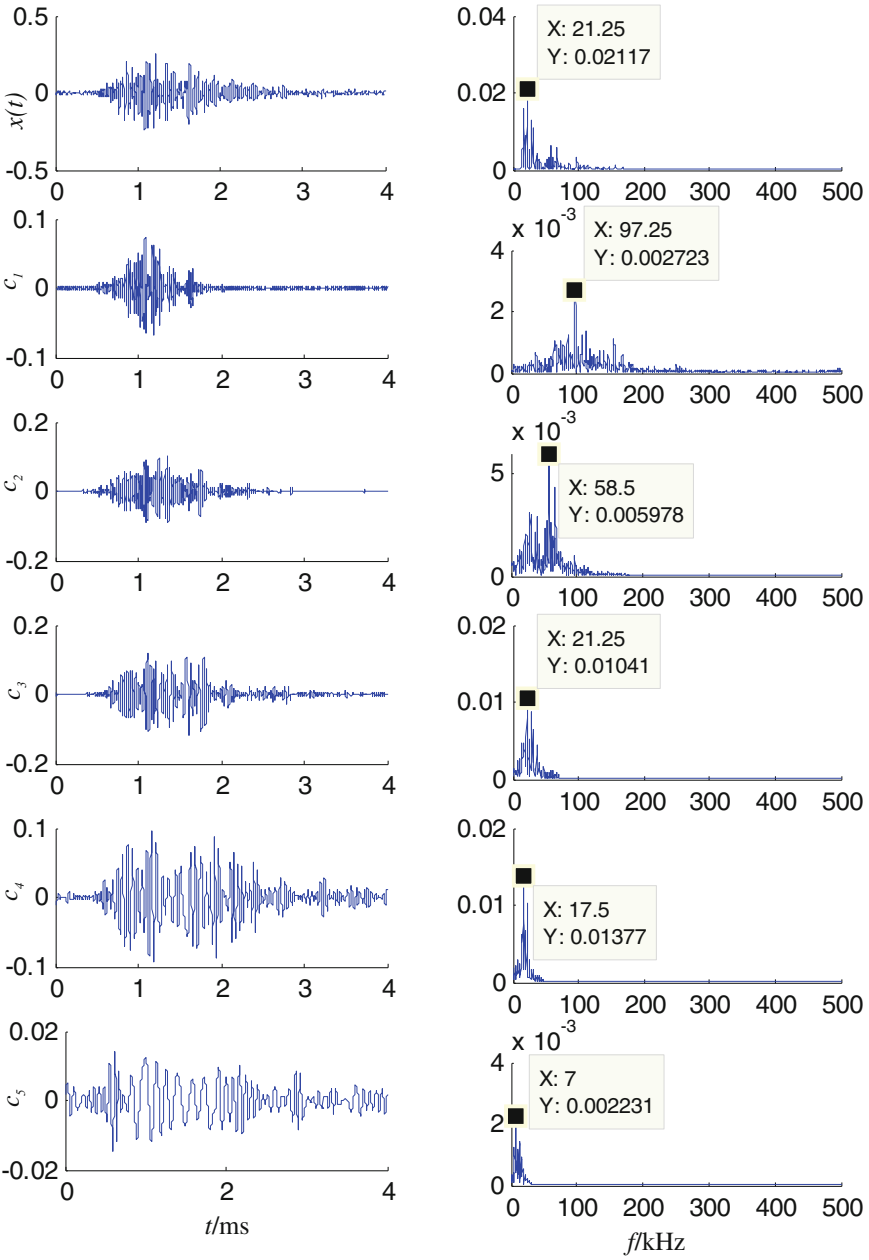


Fig. 4 The waveform, EMD components and spectra of rubbing AE

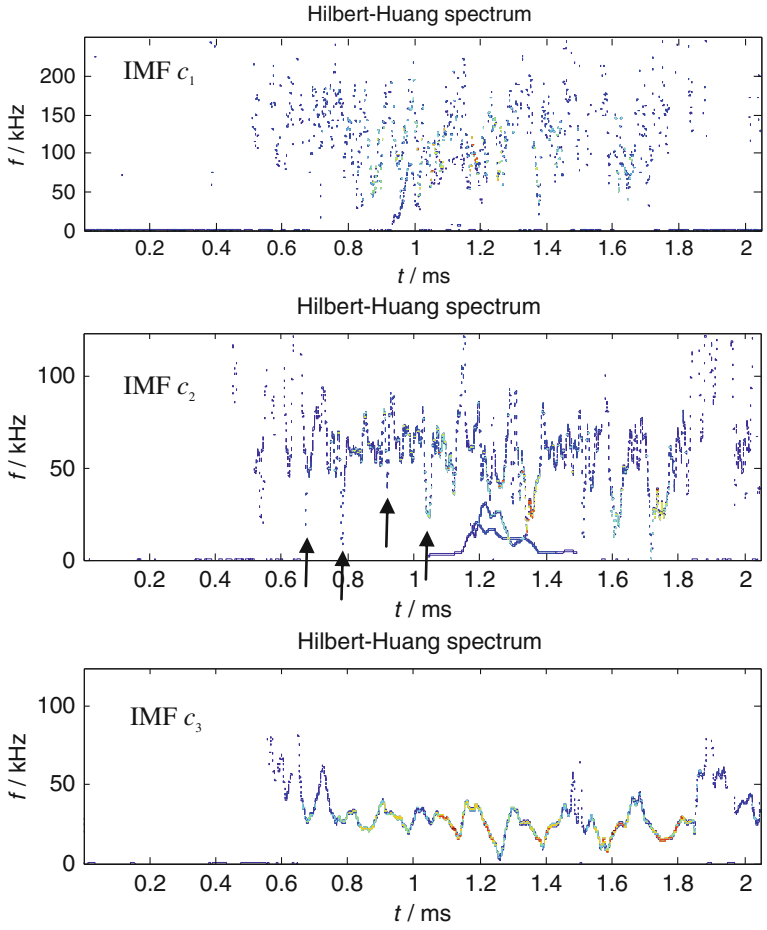


Fig. 5 The Hilbert–Huang spectra of IMF component

the propagating velocity of AE signal in the shaft is 500 mm/0.1 ms = 5000 m/s, which is very close to the longitudinal wave propagation velocity

$$v = \sqrt{\frac{E}{\rho}} = \sqrt{\frac{2.11e11}{7.85e3}} = 5184.5\,\text{m/s} \tag{12}$$

However, when the second and more rub-impacts occur, the AE signal is coupled and aliasing. This leads to the complexity of the AE signal, and it is difficult to analyze the AE signal in details.

5 Conclusions

In this paper, a special stator is designed to obtain the AE signal of rubbing fault. And the Hilbert–Huang Transform is applied to analyze the AE signal. The results show that the IMF component with peak frequency 58.5 k can be used to analyze the rubbing occurrence and the propagating characteristics in the shaft, which is helpful to diagnose the rubbing fault.

Acknowledgments This research is supported financially by National Natural Science Foundation of China (Grant No. 51175279) and Beijing Natural Science Foundation (Grant No. 3112013).

References

1. Muszynska A (1989) Rotor-to-stationary element rub-related vibration phenomena in rotating machinery—literature survey. Shock Vib Digest 21:3–11
2. Chu F, Zhang Z (1998) Bifurcation and chaos in a rub-impact Jeffcott rotor system. J Sound Vib 210:1–18
3. Chu F, Lu W (2005) Experimental observation of nonlinear vibrations in a rub-impact rotor system. J Sound Vib 283:621–643
4. Mathews JR (1983) Acoustic emission. Gordon and Breach Science Publishers Inc., New York
5. Li C, Li SY (1995) Acoustic emission analysis for bearing condition monitoring. Wear 185:67–74
6. Toutountzakis T, Tan CK, Mba D (2005) Application of acoustic emission to seeded gear fault detection. NDT&E Int 38:27–36
7. Sato I (1990) Rotating machinery diagnosis with acoustic emission techniques. Electr Eng, Jpn 110:115–127
8. Wang Q, Chu F (2001) Experimental determination of the rubbing location by means of acoustic emission and wavelet transform. J Sound Vib 248:91–103
9. Hall LD, Mba D (2004) Acoustic emissions diagnosis of rotor-stator rubs using the KS statistic. Mech Syst Signal Process 18:849–868
10. Huang NE, Shen Z et al (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc Roy Soc Lond 454:903–995

Use of Condition Monitoring in the Proactive Maintenance Strategy

Stanislaw Radkowski and Marcin Jasinski

Abstract The objectives of the presented paper are to better understand the mechanism of failure development in the dynamic mechanical systems. It is important to consider carefully the present changes in technical conditions of elements of the system when planning the requirements to a proactive risk management's strategy. While a dynamical system traditionally is modeled by structural decomposition, in the paper the dynamic behavior of system is modeled by decomposition of the behavior flow into events and errors accompanying occurrence of failure. From this point of view the nonlinear model of failure occurrence is analyzed to realizing long term benefits of a proactive maintenance strategy. The models describe the effect on the evolution at the process variables for each hypothesized fault failure. In the paper it is also analyzed the impact of nonlinearity of the sub-section on the behavior of the whole system.

1 Introduction

Proactive Maintenance Strategy (PMS) is discipline consisting of means and methods to the actual life cycle conditions to determine the advent of failure and mitigate system risks. The system incorporate PMS for number of reasons as: life cycle cost reduction, failure avoidance, future design improvement of system and its availability. When PMS is effectively implemented it means that safety, reliability and availability of system will be improved and the overall maintenance cost will be reduced.

S. Radkowski (✉)
Automotive and Heavy Machinery Engineering, Warsaw University of Technology, Narbutta 84, 02-524 Warsaw, Poland
e-mail: koch@zib.de

M. Jasinski
Institute of Vehicles, Warsaw University of Technology, Narbutta 84, 02-524 Warsaw, Poland
e-mail: jachuu@simr.pw.edu.pl

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_52

Development of mechantronic systems, especially the development of measurements and analysis of dynamic quantities, resulted in a situation in which a constructor is able to account for a product's evolution, caused by wear and tear processes during a product's operations, already during the design phase. Effective use of this knowledge will in many cases decide about the adopted operational strategy, the level and the extent of diagnostic resources involved, the method of achieving the desired safety level during each phase of a product's life, especially during the maintenance and repair phases.

As a result of finalization of a conceptual, construction or implementation designs, we reach, step by step, various variants which are subjected to valuation based on specified technical and economic criteria which enable objective decisions to be taken.

The issue of determining the criterion for selecting the relevant methods and means of diagnosis continues to remain an unsolved problem in such an approach. Let us note that the right selection of diagnostic procedures is the decisive factor in shaping the ease of diagnosis of the designed product.

The value of diagnostic information can be expressed in the form of a measure which accounts for change of the decision-makers' efficiency. In other words, the ability to provide information of relevant quality resulted in lower uncertainty as regards the right action to be taken and hence it enabled the right decisions to be made.

Let us note that while assessing the utility value of the information provided at a product's design phase, we deal not so much with the volume of information but with the impact it has on change of a decision maker's efficiency in respect of maintenance activity.

2 Proactive Maintenance

Idea of proactive maintenance system algorithm was presented in literature [3, 8, 9] (Fig. 1).

Let us note that estimation and modeling of the degradation process is one of the most effective methods of defect development anticipation and maintenance of system operation in terms of nominal parameters. In reality such an approach denotes compilation of several conventional methods of forecasting – probabilistic behavioral models and event models in particular. Probabilistic behavior and degradation models enable analysis of the type and extent of uncertainty which conditions forecasting reliability. Event models are a kind of a combination between the contemplated models and the actual system and they make up the basis for constructing and analyzing causal models which enable assessment of degradation and determination of the optimum scenario of maintenance-and-repair work.

As a result, the process of defect origination can be analyzed and included in new structural solutions while maintenance of existing machines can be corrected accordingly. In this latter case the main task involves diagnosis of the defect initiation period (Fig. 2).

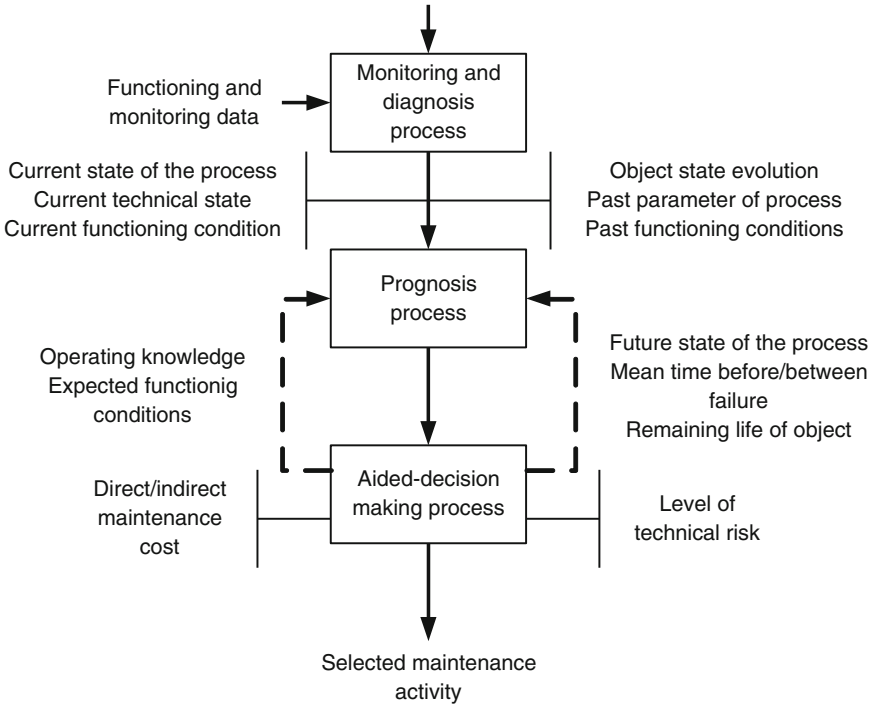


Fig. 1 Architecture of the proactive maintenance system

From the point of view of relevance of diagnostic information in the contemplated pro-active maintenance strategy, it is the adoption of a relevant diagnostic-and-prognostic model that is one of the basic problems to be solved. Publications on the topic discuss a whole spectrum of proposed models: starting from the models which enable qualitative description and understanding of the analyzed processes only, though the models which give insight into the general trend of diagnostic parameters' changes, to the models which have the form of a virtual laboratory which simulates actual maintenance processes.

A separate group consists of models which enable examination of the potential behavior of the monitored object on the basis of a relevantly selected vector of diagnostic parameters.

Assessment of an analyzed diagnostic parameter or of a method of diagnosis should be carried out while using relevant criteria of usefulness of diagnostic information in valuation and decision-making processes. It is the systematic analysis of diagnostic goals, accounting for all the essential features of the diagnosed system and enabling comprehensive assessment of alternative solutions that should decide about the adoption of a specific solution.

The basis of reliability and rationality of a diagnostic system is made up by the correctly formulated and solved task of diagnostic parameters' selection.

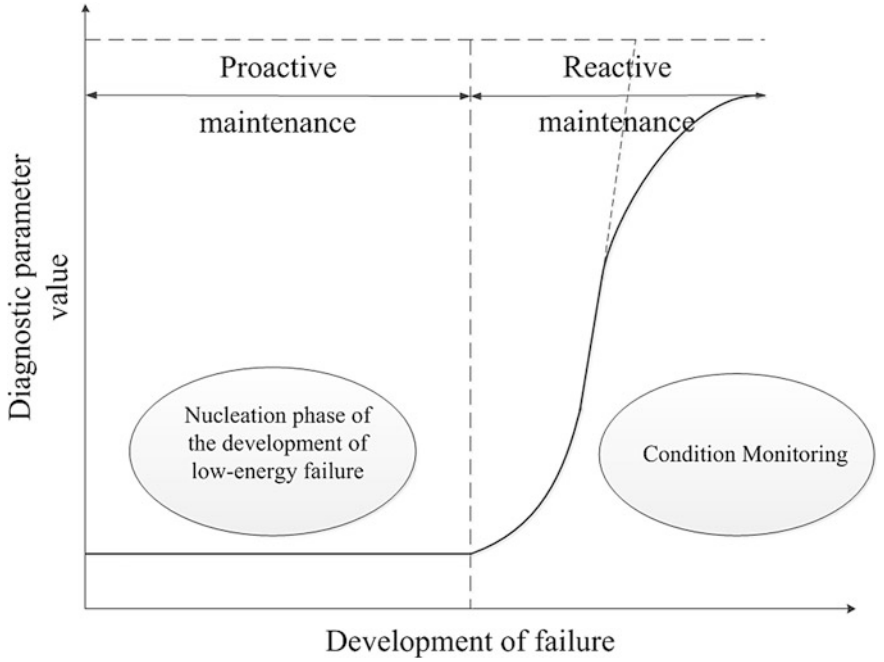


Fig. 2 Diagnostic tasks in proactive operation

Let us note that while implementing a proactive maintenance strategy the examined system may not reviewed and modeled only while using the structural decomposition rules. We should account for the dynamic behavior of the system. Accordingly, the sub-systems and units should be decomposed while accounting for the behavioral influence of errors, events and decisions. At the same time this type of decomposition should serve as the basis for indentifying the activity and the involvement of respective elements in fulfillment of the tasks as well as preservation of the quality of their fulfillment. It is also for that reason that description of the tasks should include not only the sequences of activities but also the possibility of selecting a relevant solution and the degree in which the selection depends on the quality of fulfillment of functional tasks, level of threat and level of potential minimization of uncertainty regarding the effects of the decisions taken in the system itself and with regard to the system.

3 Modeling of Vibroacoustic Signal of Low Energy Failures

In accordance with the assumption of the need to analyze the energy dissipation process, for the problem of diagnosis of origins of toothed wheels defects that was presented here, the author proposes a model, the basis of which is the relationship

describing power of the friction caused by mating of a pair of toothed wheels along the path of contact:

$$N_T = \int F_R v_g \frac{dt}{t} = \int \frac{v_g^2}{|v_g|} \lambda_R F \frac{dt}{t} \tag{1}$$

Where according to [11]:

$$F \approx \frac{M_R}{r_R}; \lambda_R = \frac{\mu}{1 + \frac{v_g}{|v_g|} \mu ctg\kappa} \tag{2}$$

where:

μ Friction coefficient,
λ_R Friction factor refereed to path of contact,
v_g Sliding velocity,
F_R Friction force

$$\kappa = \frac{\pi}{2} - \alpha_y - \text{ angle of shift} \tag{3}$$

The angle of shift can be defined with the use of the following relationship:

$$\kappa = \frac{\pi}{2} - \arccos \frac{r_b}{r_y} \tag{4}$$

The thus constructed model enables a separate analysis of disturbances of the sliding velocity the force working between the teeth, and the analysis of the resultant structure of the vibroacoustic signal's spectrum. At the same time, the introduction, after Roth, of the dependence of friction coefficient upon the location on the path of contact enables us to take into account the qualitative change of the disturbance in the point of rolling contact, and as a result to exhibit the significance of the 1st harmonic of meshing in the diagnosis of the general condition of a toothed gear [1].

On the other hand the proposal of such modeling of the vibroacoustic signal's disturbances has the traits of a more general solution of the problem. To stress this fact, let us assume that the signal generated according to relationship (1) is transmitted to the measuring point by a linear transmission channel which corresponds to the process model with linear elements and multiplying elements, which can be defined by means of a Volterra series [12].

After including the linear part of the signal, for which the dynamic part of the system will be responsible in this case, and the earlier defined non-linear part of the process (Fig. 3), we will obtain the following relationship in accordance with the procedure proposed by Eykhoff [2]:

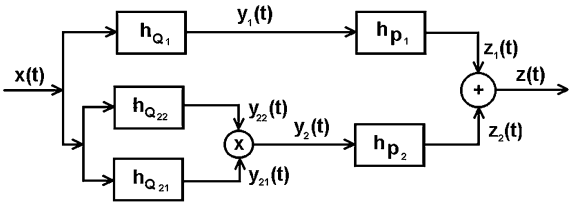


Fig. 3 Diagram of generation of linear and non-linear parts of the process

$$z(t) = \int_0^t h_1(\tau_1)x(t-\tau_1)d\tau_1 + \int_0^t \int_0^t h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 \tag{5}$$

where:

$$h_1(\tau_1) = h_{Q_1}(\tau_1) * h_{P_1}(\tau_1) \tag{6}$$

$$h_2(\tau_1,\tau_2) = h_{Q_{22}}(\tau_1-\tau_3)\cdot h_{Q_{21}}(\tau_2-\tau_3) * h_{P_2}(\tau_3) \tag{7}$$

As was showed in [10] it is possible to use the simplified model, which reduces the number of coefficients required for a Volterra series representation. The second-order Volterra kernel will be reduced to:

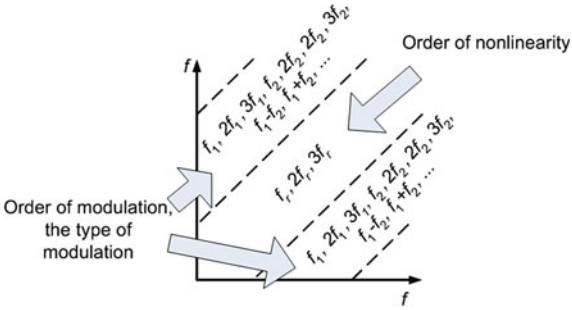
$$h_2(\tau_1,\tau_2) = h_1(\tau_1)\cdot h_1(\tau_2) \tag{8}$$

Knowing the linear impulse response and the input, we can get the second order Volterra operator:

$$H_2(x(t)) = \left[\int_{-\infty}^{\infty} h_2'(\tau)\cdot x_1(t-\tau)\mathrm{d}T\right]^2 \tag{9}$$

Its mean, we are interested in diagonal of bispectral plane (Fig. 4).

Fig. 4 Bispectral plane



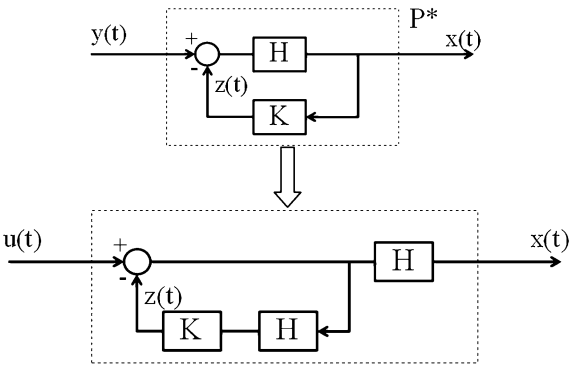


Fig. 5 Feedback system

This model was tested to descript the phenomenon of failure nucleation.

A nonlinear feedback system due to failure development was presented at Fig. 5. Where:

$$P_1 = H_1 \cdot R_1 \tag{10}$$

$$P_{21} = H_1 \cdot R_2 + H_2 \cdot R_1 \tag{11}$$

$$R_1 = \frac{1}{1 + H_1 \cdot K_1}; \quad K_1 = \frac{1}{H_1(s)} \tag{12}$$

$$R_2 = -R_1 \cdot [H_2 \cdot K_1 + H_1 \cdot K_2] \cdot R_1 \tag{13}$$

$$K_2 = -K_1 \cdot H_2 \cdot K_1 \tag{14}$$

$$X(t) = P[y(t)] = \sum_{k=1}^{n} P_k[y(t)] \tag{15}$$

For second order Volterra series (15) we got [12]:

$$X(t) = P_1[y(t)] + P_2[y(t)] \tag{16}$$

4 Results of Laboratory Experiments

The experiment was conducted at the FZG back to back test-bed [6, 7]. The test-bed consists of two toothed gears operating in a revolving power setup and it enables examination of both toothed wheels as well as gear lubricants.

The shaft connecting the pinions is divided, which enables rotating one of its sections versus the other and thus introducing relevant meshing forces. Strain gauges are affixed to the shaft and they enable measuring the torque. Wheels with straight teeth are installed in the examined gear, while wheels with helical teeth are installed in the closing gear. Thanks to such a set up it was the examined toothed

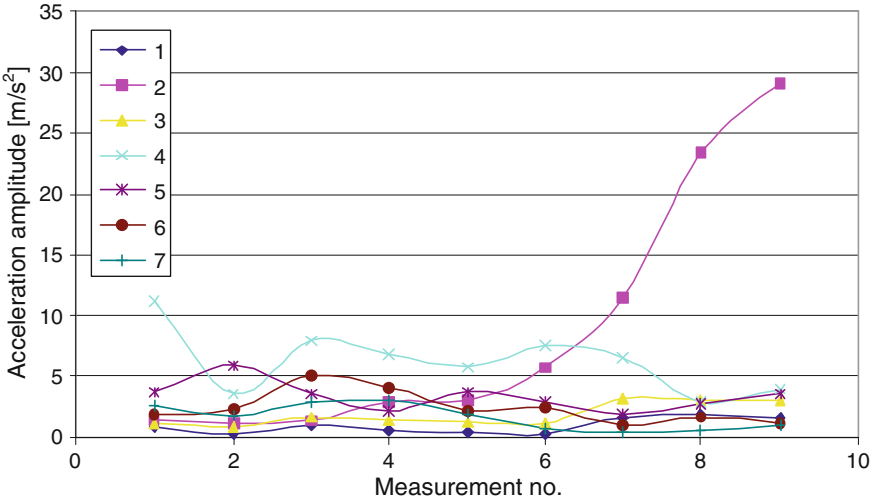


Fig. 6 The changes of subsequent mesh harmonics (no. 1–7)

gear that was subject to defect-development during the experiment. Parameters of the test-bed:

- Maximum tensioning torque 1,200 Nm (or 1,500 Nm for shafts with bigger torsional rigidity):
- Motor speed: 1,460 rpm;
- Gear ratio in both toothed gears: 1.296;
- Module of test specimen wheels and counter-test specimen wheels 4 mm;
- Number of teeth in test specimen wheels: 27;
- Number of teeth in counter-test specimen wheels: 35;
- Axle base for wheels: 125 mm.

Toothed wheels made of 20H2N4A carburized steel, hardened to 60 HRC hardness were used for the research. They were subjected to accelerated fatigue test.

Figure 6 present the changes of subsequent mesh harmonics of a vibroacoustic signal registered on the toothed gear's casing during the whole experiment.

The changes which accompany the subsequent phases of development of fatigue-related defects are observable in a bispectrum [5]. Particularly interesting results have been obtained for a diagonal bispectral measure, for a maximum bispectral measure and for measure created from vector of maximum values of triangular matrix separated from bispectrum matrix by removing main diagonal of this matrix [4]. As a result, the phase reactions defined by the dominant non-linear effect become blurred. The results point to high sensitivity of bispectral measures to changes of the signal's frequency structure and to the possibility of using these relations while constructing models of development of degradation-and-fatigue-related processes which are required while creating the procedures of proactive maintenance strategies.

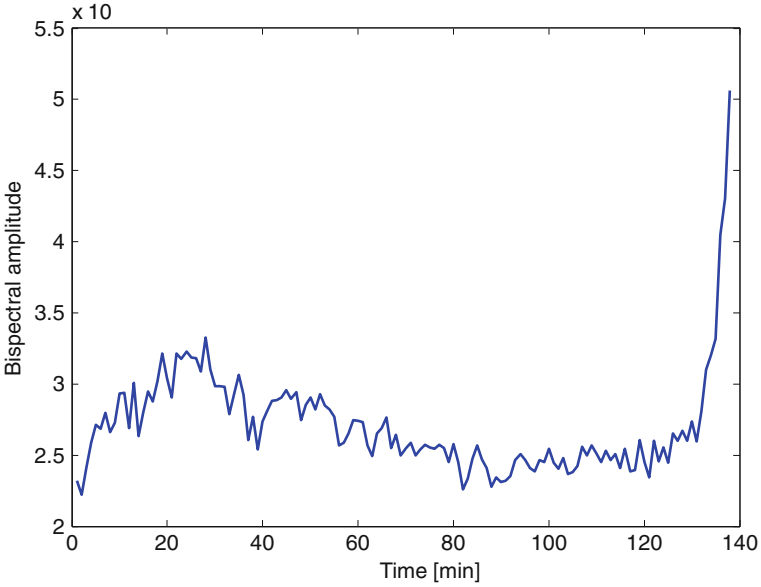


Fig. 7 Integral of bispectral noise from bispectral residual maximum diagrams, full investigation of wheel no. 7

Next step was to create a new measure which is able to predict the moment of fatigue tooth crack. Integral of bispectral noise from bispectral maximum diagrams and integral of bispectral noise from bispectral residual maximum diagrams (Fig. 7) were calculated with maximum level 0.5E8 (everything higher than maximum level was equalize to this maximum level) for full life time of this wheel. At Fig. 7 we can see that calculated derivative of this diagrams (applying a smoothed curve) we can build effective and sensitive diagnostic parameter of quality changes of fatigue process of toothed wheel damage.

5 Conclusion

While planning a proactive maintenance strategy, we in reality analyze the system of proactive risk management in a system. This means the necessity of understanding how threats are generated at all levels of the system and having the ability to make relevant decisions.

It is indispensable to understand the necessity of identification of the decision-makers' information-related needs, both as regards the actual situation as well as identification of the aspects which are essential to improve the quality of fulfillment of tasks and achievement of goals. As a result, proactive risk management in a transport system calls for the following activities to be undertaken:

- analysis of normal activity of the system while examining the properties having the biggest influence on the system's behavior in the future;
- examination of interactions and changes in a system's structures from the point of view the control systems' theory;
- monitoring the possibilities of information flow, access to relevant information sources, selection of relevant methods of detection and identification as well aggregation of information from the point of view of implementation of the tasks associated with technical risk management;
- formulation of recommendations of actions aimed at improving the quality of a system's operation, minimize the number of errors and reduce the hypothetical consequences of undesirable events.
- To recapitulate, the proactive maintenance strategy is not an attempt to eliminate the reasons of errors, especially human errors, but it rather focuses on designing the strategies of use of machines in a way enabling:
- identification of the limits of safe system functioning while making these limits legible and visible in the decision-making process;
- minimization of influence of the factors which are conducive to the decisions to exceed the assumed limits of safe system operation.

References

1. Bartelmus W (1992) Vibration condition monitoring of gearboxes. Mach Vib 1:178–189
2. Eykhoff P (1980) Identification in dynamic systems (in polish). PWN, Warsaw
3. Han T, Yank BS (2006) Development of an e-maintenance system integrating advanced techniques. Comput Ind 53:569–580
4. Jasinski M, Radkowski S (2011) Use of the higher spectra in the low-amplitude fatigue testing. Mech Syst Signal Process 25(2):704–716. doi:10.1016/ymssp.2010.06.001,2011
5. Jasinski M, Radkowski S (2010) Use of bispectral-based fault detection method in the vibroacoustic diagnosis of the gearbox. Eng Asset Lifecycle Manag 19:651–660. doi:10.1007/978-0-85729-320-6_76
6. Maczak J (2013) Local meshing plane analysis as a source of information about the gear quality. Mech Syst Signal Process 38:154–164. doi:10.1016/j.ymssp.2012.09.012
7. Maczak J. Radkowski S., (2001) Use of factorial simulation experiment in gearbox vibroacoustic diagnostics. In: Proceedings of the 14th international congress of condition monitoring and diagnostic engineering management, Manchester, UK
8. Muller A, Suhner MC, Iung B (2008) Formalisation of a new prognosis model for supporting pro-active maintenance implementation on industrial system. Reliab Eng Syst Saf 93:234–253
9. Muller A., Suhner M.C., Iung B. (2007) Maintenance alternative integration to prognosis process engineering. J Qual Maint Eng (JQME) 2
10. Novák A (2007) Identification of Nonlinear Systems: Volterra Series Simplification. Acta Polytechnica 47(4–5):72–75
11. Roth K (1989) Zahnradtechnik. Springer, Berlin
12. Schetzen M (1980) The Volterra and Wiener Theories of Nonlinear Systems. Wiley, New York

Diagnostic Model of Hysteresis for Condition Monitoring of Large Construction Structures

Szymon Gontarz and Stanisław Radkowski

Abstract The conception of diagnostics, based on passive magnetic field measurements is being presented. It was shown that the transformation of the signal to the appropriate form of hysteresis, creates new potential for identifying changes in signal parameters that correspond to specific physical phenomena. Model supported analysis of changes in the magnetic field, made it possible to obtain diagnostic parameters characterizing magnetomechanical hysteresis, which can be used in determining the state of the material strain on the object structure subjected to variable loads. Suitable model of hysteresis was elaborated and conditions in which it could be applied were stated. Obtained by authors diagnostic parameter confirm the validity of the hysteresis model used for this case. The chapter concludes that state of polarization of the magnetic structure (the distribution of magnetization) is carrier of diagnostic information about the level of effort and the progressive degradation of the material structure. It is possible to obtain the diagnostic information through a remote, non-contact measurement of magnetic field near the test structure.

1 Introduction

Materials that run a high risk of damage due to material fatigue, exceeding maximum stress or plastic deformations have magnetic properties, which allowed to develop a group of magnetic methods in technical diagnostics. Currently, a group of passive diagnostic methods is developing dynamically in parallel to active diagnostic methods [1, 2]. The group of passive diagnostic methods has all the advantages of active diagnostic methods group, and additionally it does not require

S. Gontarz (✉) · S. Radkowski
Integrated Laboratory of the Mechatronics System of Vehicles and Construction Machinery, Warsaw University of Technology, Narbutta 84, 02-524 Warsaw, Poland
e-mail: sgontarz@simr.pw.edu.pl

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_53

the application of artificial source of magnetic field that needs sophisticated and expensive equipment.

Owing to this quality, passive methods offer a possibility to be applied not only in temporary diagnostic tests, but also on a permanent basis e.g. in condition monitoring.

The utility of the methods can be observed in SHM system. In addition, passive magnetic methods work primarily on the basis of interaction with the Earth's magnetic field. Although it has low strength values, when the stress factor in materials with magnetic properties has been changed, the material changes into magnetic state that may also be analyzed from a distance. Generally it may be stated that the characteristic properties of passive diagnostic methods described above may be used for large construction structures monitoring.

2 Application of Earth's Magnetic Field in Diagnosing Stress Condition

From the perspective of construction structures diagnostics it is evident that all types of steel constructions will have magnetic properties. It was noticed that changes in ferromagnetic object's own magnetic field may occur due to geometric discontinuities of the material, e.g. ruptures or high concentration of dislocations in the structure of material, but they may also be caused by the change of stress conditions. The following phenomena take place in the presence of the Earth's magnetic field and they may be interrelated with magnetic state of an object (magnetization). Taking into account magnetic properties of steel construction materials and values of the Earth's natural magnetic field, the resultant of volume magnetic susceptibility M_{vol} may be omitted. Therefore, the current degree of an object's (material) magnetization can be expressed as the resultant of reversible directional magnetization process and irreversible magnetization:

$$M(\sigma, t) = M_{dir}(\sigma, t) + M_{irr}(\sigma = 0, t) \tag{1}$$

Hence, own magnetic field of the object H measured by e.g. a magnetometer depends on the object's magnetization and its distribution in the space. Additionally, taking into account scientifically recognized magneto-mechanical phenomena [3], namely the Villary effect, the Matteuci effect, the Naganka-Honda effect or the phenomenon of austenite transforming into martensite in fatigue loads, it turns out that the change of stress degree in an object will be reflected by a corresponding change of magnetic field [4, 5].

Let us then verify telediagnostic possibilities in case of analyzing magnetoelastic effects in stretching a steel specimen in laboratory conditions. A specimen of a circular cross section (fi = 6 mm) was designed and in theory, the distribution of its own magnetic field around this specimen should be regular. The following type of steel was used to design the specimen: C45 (PN-EN 10083-2:1999). In order to

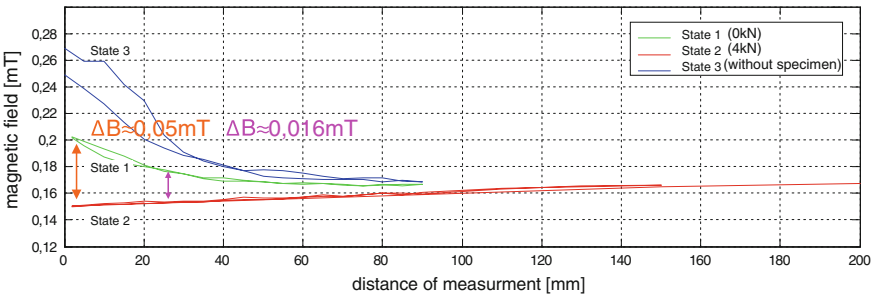


Fig. 1 Distribution of own magnetic field sample in a space

identify the Villary effect, the generated signal of local magnetic field's strength changes was registered without additional artificially generated magnetic field sources, in standard laboratory conditions with regular work of electrical installation. The specimen was loaded with a force that guaranteed specimen's deformations in elastic range. In order to perform the measurement, a magnetometer type fluxgate APS536 was used.

The distribution of magnetic field in space was measured for three states: measurement at a laboratory stand without the specimen, measurement at a laboratory stand with the specimen and measurement of the specimen loaded with 4 kN force. The results were registered by a magnetometer that changed its position towards the specimen depending on the state. The following graphs present the results (Fig. 1).

The graphs present several interesting effects. When no specimen is present, a certain value of the background magnetic field strength was registered and it was certainly formed by a universal testing machine. The distribution of this field in the distance function definitely resembles the theoretical relation. The shift of initial and final values of magnetic field for the same distance from the stand results from the lack of accuracy in the movement of the sensor towards the stand (the lack of the sensor caused problems with defining an adequate point of reference). The applied specimen brought about an interaction with the universal testing machine's own magnetic field and a change of the received value. The distribution of such resultant field in the space practically remained unchanged. The fact that the specimen was loaded and magneto-elastic effects were generated drastically changed the view of magnetic field. The effect was so strong that it changed the sense of a vector of the magnetic field's strength and additionally, it was visible at a distance of about 80 mm. Above this distance all three lines are going in the direction of a common asymptote, which proves a correct interpretation of the object's own magnetic field.

Telediagnostic possibilities in case of analyzing magnetoelastic effects in stretching a steel specimen in laboratory conditions were also verified in [6]. The results obtained allow us to state that the Earth's magnetic field influence on a loaded object may be observed from a certain distance. Additionally, it may be

observed that for the same degree of stress showing magneto-elastic effects, different values of own magnetic field signals may be registered. This may occur due to natural reasons, e.g.—geographical location in case of local anomalies caused by the structure of the crust, or for artificial reasons, e.g. neighbouring objects that are sources of strong magnetic field. It has to be taken into account though, that as the distance increases, the risk of receiving magnetic field from another source arises. This may be a problematic situation, however the application of e.g. differential measurement with at least two sensors should solve the problem. Another observed fact is that the distance has a direct influence on quantity change, whereas quality change in the object's magnetic field may still be visible and detectable independently from the distance.

3 Hysteresis in Passive Magnetic Method

Taking into account the possibility to identify magneto-elastic effects from a distance one should consider how to measure, analyze and interpret the results. We can notice that the phenomenon of hysteresis depends on various factors proper for the problem analyzed. The goal of the analysis proposed in the present chapter is to present the possibilities of applying hysteresis model for passive diagnostic method. Based on the analysis performed so far one can propose a hypothesis saying that an adequate model of hysteresis will describe the change of the registered magnetic field's signal depending on the load and deformation of the construction's material.

Hysteresis may be generalized to a system generating an input signal and a system causing the input transformation in such a way that it is possible to register hysteresis. The input signal is a signal of magnetic field strength. The output signal is the registered signal that has gone through the transformation system with given parameters. An adequate adjustment of the system transforming the signal and related with magnetic effects should provide access to information about the operation of the system generating the signal.

Analyzed phenomena that caused registered change of own magnetic field [6] are of a magneto-static nature. Theoretically this fact excludes the possibility of hysteresis because the prerequisite for its occurrence is an input signal of an oscillatory character—like in case of typical magnetic hysteresis. Let us mention, however, the idea of the changes of an object's own magnetic field under the influence of magneto-elastic effects and in view of the possibilities to observe the hysteresis loops. The magnetic anomalies measured stand for the change of magnetic field's induction in the presence of the assumed constant external magnetic field $H = const$. This external field is identified with the Earth's magnetic field that —for the purposes of the present chapter—may be regarded as constant, according to the results presented above. If we use an additional, already verified hypothesis saying that magnetic permeability of a given object's material depend on the degree of its stresses, we may draw the conclusion that the shape of hysteresis will be obtained during cyclic changes of the examined object's loads. Introducing the

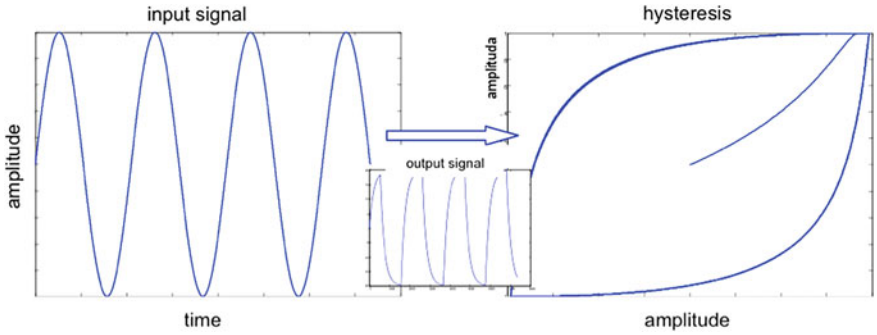


Fig. 2 Example of hysteresis loop forming after dynamic input signal

cycle of dynamic loads as an input signal into a hypothetical model of hysteresis, we will get an output signal that represents forming of a hysteresis loop at the level of input/output (Fig. 2).

It may be observed that after each full cycle of loading is completed a hysteresis loop appears. The introduction of dynamic load that will definitely be the case when real objects are analyzed, exposed the occurrence of hysteresis process. However, the above mentioned situation is more complex as other conditions (factors, relationships) that result from the specificity of own magnetic field's changes depending on the loads must be taken into account.

Based on the analysis presented above it may be concluded that the characteristic behaviour of a ferromagnetic material, namely its non-linear magnetization under the influence of external alternating magnetic field (classic magnetic hysteresis) may also occur in the presence of constant magnetic field (even a weak one). However, this is only possible under the condition of the change of ferromagnetic stress state.

The hysteresis area in case of periodic change of the magnetic induction in one cycle and one material volume depends on the frequency of changed and may be described by the following formula:

$$\Delta E_d = C_0 + C_1\omega + C_2\omega^{\frac{1}{2}} \tag{2}$$

where C_0, C_1 and C_2 are material constants, whereas ω is the frequency of input signal changes.

However, our particular case allows to omit C_1 and C_2 factors for very low velocities of input signal changes (input by load factor). Additionally, the induction of eddy currents should also be expected, however this effect may also be omitted for low frequency of inputs.

Based on the analysis presented above it may be concluded that the characteristic behaviour of a ferromagnetic material, namely its non-linear magnetization under the influence of external alternating magnetic field (classic magnetic hysteresis) may also occur in the presence of constant magnetic field (even a weak one).

However, this is only possible under the condition of the change of ferromagnetic stress state.

Let us attempt at defining the nature of magneto-mechanical hysteresis. We may generally state that certain phenomena of a magneto-static character occur under the influence of mechanical forces. An identified hysteresis is related to a specific character of ferromagnetic material magnetization, in constant Earth's magnetic field, with alternating loads: $B = \mu(\sigma) \cdot (H + M)$, H, M = "*const*"$\rightarrow$ $B = f(\sigma)$ because μ = f (F). The shape of hysteresis depends strictly on the character of the material's load that defines characteristic points of the loop. Performing a constant loading of a ferromagnetic from the value σ_{min} to σ_{max} and back from the value σ_{max} to σ_{min}, we obtain a closed loop (proper for a given material). If the hysteresis does not reach maximum values for a given material in given conditions, then the shape of the loop will resemble an elongated ellipse (a situation when final points of the hysteresis loop will not be part of magnetic saturation area). However, each change of the value of stress in the material in form of the elastic deformation will bring about the changes of the shape and the field and will tend to achieve a classic shape of hysteresis. The increase of the magnetic field induction current will be expressed by the following formula: $d\Phi = d(BS)$, where B stands for the induction of ferromagnetic material's magnetic field that depends in a hysteretic way on the alternating degrees of stresses in the material. Considering the character of the hysteresis loop formed, let us assume that $f = \frac{1}{T}$. Accordingly, when $f \rightarrow 0$, then $T \rightarrow \infty$. The examination and the analysis of the phenomenon show that when $f \rightarrow 0$ the loop of hysteresis does not disappear, because $H_T(u) \rightarrow H_\infty(u)$ for $T \rightarrow \infty$, hence $H_\infty(u)$ has (u, y_1), (u, y_2) and $y_1 \neq y_2$, which confirms the definition of the system with hysteresis. This means that the loop of hysteresis was formed as a result of the system's properties with the assumption of a periodic input signal of C^0 class.

4 Experimental Verification of Hysteresis

Development of a full diagnostic model, while using and modifying the general hysteresis model, requires an experiment which verifies the described theoretical considerations. An experiment was planned which registered and analyzed the distributions of own magnetic fields of objects made of magnetic materials. Measurements concerned dynamic responses of the examined samples. The measured anomalies are intended to present change of magnetization which reflects the operation of major stresses, generated by the loads which were in turn caused by degradation or operation of a structure. The observed signals, having the form of change of own magnetic field, will emerge during actual deformations, leading to structural changes in the examined object.

A measurement track, consisting of an APS536 FluxGate-type magnetometer (Fig. 3), a National Instruments data acquisition card and a measuring computer,

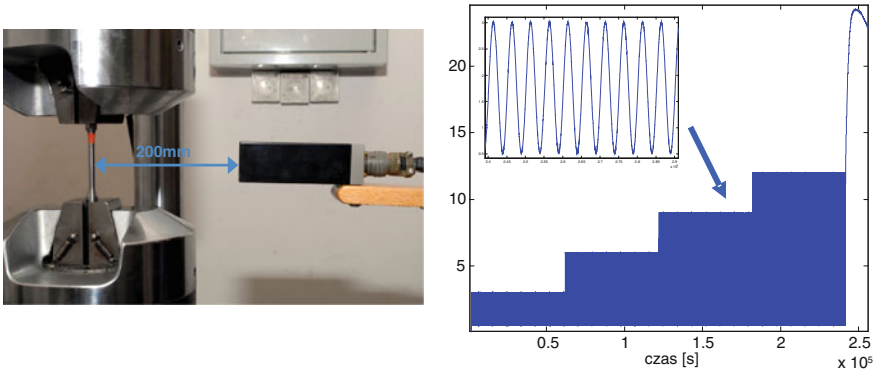


Fig. 3 *Left side* A view of the testing machine's shoe, the sample and the magnetometer *Right side* The cycle of applied loads

was used to carry out the experiment. The examined objects were cylindrical samples made of a material with ferromagnetic properties, namely the C45 steel, which are visible on the photo (Fig. 3). The shape and the dimensions of the sample were designed from the point of view of using them in a machine with a specific type of a shoe and while accounting for the dynamic nature of the loads to which the sample was exposed. Load could be applied to the sample thanks to using a testing machine manufactured by MTS Systems Corporation—Material Testing System MTS 810 (Fig. 3).

The program of the tests accounted for the impact that the sample's deformations had on the behaviour of the sample's own magnetic field under regularly applied stress in the presence of the Earth's natural magnetic field H. The initial load was 0.5 kN while the nature of the deformations was a regular (sinusoid) change of the force, from the value of the load to a predefined value of the force, increasing in steps by respectively 3, 6, 9 and 12 kN (Fig. 3). The system subjected to a load responded with an induction signal of the sample's magnetic field, measured in three directions. Apart from that, it was the growth of the sample's length as well as the value of the force exerting the load that were measured. The load-exerting force was controllable during the experiment while its change was realized for two frequencies: 0.5 and 1 Hz.

In accordance with earlier considerations, the chance of observing a magneto-mechanical hysteresis loop exists when a load (e.g. a deformation force) is presented as a function of change of a magnetic field's induction. Figure 4 shows that the characteristic loops emerge in this very domain for the subsequent load cycles. The next Fig. 4 presents selected loops which emerge in the case of an input having the frequency of 0.5 and 1 Hz, and amplitude of respectively 3, 6, 9 kN.

First, let us note that a cyclic change of load results in the growth of the mean value of magnetization of a sample, which is visible in the examined case as a growing level of emergence of the loop (Fig. 4). As it turns out, it is this very phenomenon, registered and identified in a different way, that serves as the basis for

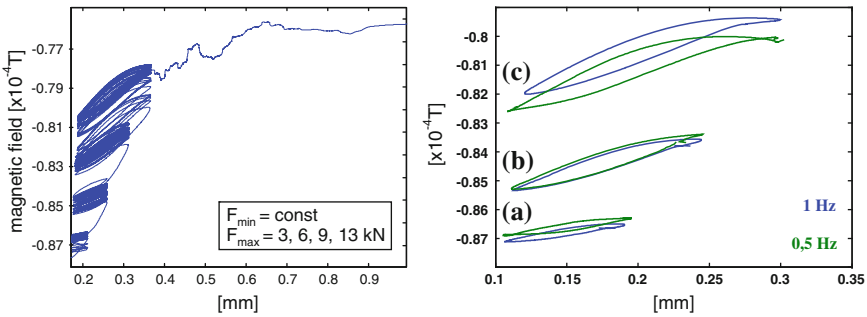


Fig. 4 Left side: Change of induction of the magnetic field in the function of deformation of the samples for a dynamic load test. Right side: Hysteresis loops for different frequency exciting force at different loads: **a** 3 kN, **b** 6 kN, **c** 9 kN

operation of the most popular passive magnetic method, namely the magnetic metal memory method (MMM). Coming back to the experiment, the analysis of the hysteresis loop seems more interesting here. In this case it emerges in the area flexible deformations. In the situation of the magnetic field's induction and the sample's deformation we observe how the area of the loop's surface changes depending on the cyclic exciting force. For small loads (3 kN) there emerges a loop which increases and takes the classical shape of hysteresis as the load increases. In accordance with the nature of magneto-elastic phenomena related to elastic deformation, growth of the size of the area of the hysteresis is proportionate to the load-exerting force and is approximately linear. Each emerging loop requires a full cycle of the exciting force to be completed.

The observed condition of the sample's own magnetic field has magnetostatic nature (similarly as the nature of Villari's phenomenon), however the oscillation of the input signal (e.g. of the force or of the stretching) introduces the required "dynamics" thanks to which there emerge sufficient conditions for creation of a relevant shape of the hysteresis. Under cyclic, regular loads the hysteresis loop was emerging in the plane defined by the value of own magnetic field's induction and the value of the exciting force, while the area of its surface could have been treated as a parameter indicating the loads prevailing in the examined object.

5 Diagnostic Model of Hysteresis—Proposal and Its Verification

Currently we are observing a great progress in the development of non-linear models [7] especially the ones which are characterized by hysteretic behaviours [8–10]. However, the presented proposals are rather complex and often require additional selection algorithms and optimisation of their parameters. As an alternative, we propose our original hysteresis model for the purposes of the issues discussed in the present chapter.

The model is based on the transformation of the energy carried by the input signal. The algorithm used in the model calculates a certain equivalent of an integral that is the area under the curve. However, contrary to the integral it adds up rectangles, whose sides' lengths are the distances from the next point where the graph crosses zero to the current argument and the value of the function in the point corresponding to the argument. The sums of rectangles above the OX axis are considered as positive and the ones below the axis as negative. By definition, the input signal should have the graph of a sinusoidal signal curve, and that is why according to the assumption (paragraph 3) there is a possibility to form a characteristic hysteresis loop. The hysteresis is produced because of two reasons, firstly the sum of surfaces below the OX axis is gradually subtracted from the complete sum above the curve and secondly because of the fact that the length of one of the rectangle's sides is zeroed when crossing the zero point of the signal. The whole phenomenon is described by the graph below (Fig. 5) and the following formula:

$$\sum_{i=1}^{k} x_i \cdot y_i - \sum_{i=k}^{n} x_i \cdot y_i \tag{3}$$

This way the described model has been provided with an additional function, allowing for the model's behaviour to change from 'rate dependent' into 'rate independent', that is making the model reaction independent from variable frequency input signal. It has been achieved by applying variable step sampling in such a way that every variable period of signal has a constant number of samples for a cycle. In other words the signal sampling is carried out in dynamic way (that varies in time) in a form of linear increase of the step complying with the change of

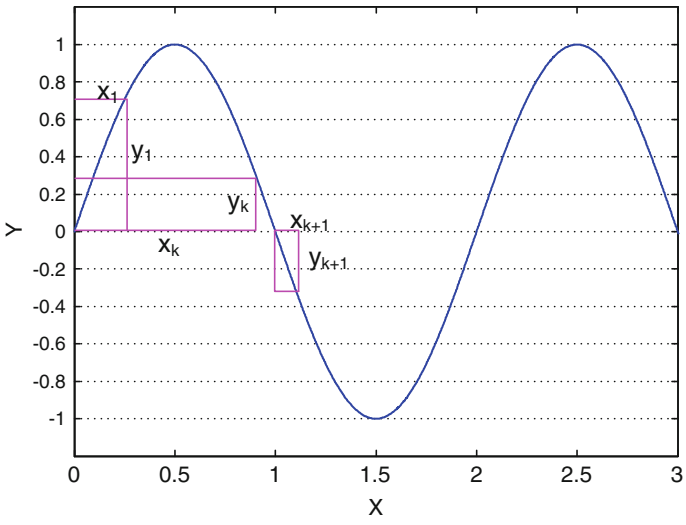


Fig. 5 Functioning of the proposed hysteresis model

the received frequency. According to the presented assumptions we can proceed to the mathematical algorithm description.

The number of samples in subsequent periods has to be constant and it equals:

$$n = \frac{T_1}{dt_1} = \frac{T_2}{dt_2} = \cdots = \frac{T_N}{dt_N} = const \tag{4}$$

hence:

$$\frac{T_1}{dt_1} = \frac{T_2}{dt_2} \;\; and \;\; \frac{T_2}{dt_2} = \frac{T_3}{dt_3} \;\; hence \;\; : \frac{T_{N-1}}{dt_{N-1}} = \frac{T_N}{dt_N} \tag{5}$$

after transforming we get:

$$dt_2 = \frac{T_2 \cdot dt_1}{T_1} \text{ and } dt_3 = \frac{T_3 \cdot dt_2}{T_2},$$

hence:

$$dt_N = \frac{T_N \cdot dt_{N-1}}{T_{N-1}} \tag{6}$$

the general form will be the following:

$$dt_i = \frac{T_i \cdot dt_{i-1}}{T_{i-1}} \tag{7}$$

which corresponds to:

$$dt_i = dt_1 \times \frac{T_2}{T_1} \times \frac{T_3}{T_2} \times \frac{T_5}{T_4} \times \frac{T_7}{T_6} \times \cdots \times \frac{T_i}{T_{i-1}} \tag{8}$$

after reducing suitable periods in formula (8) we finally receive:

$$dt_i = dt_1 \cdot \frac{T_i}{T_1} \tag{9}$$

In order to receive new arguments' vector for the input signal of the calculated lengths of the steps it is necessary to calculate the cumulated sum of elements dt_i. This way we will obtain arguments and corresponding values of the function for which the described model will create a hysteresis loop resistant to variable frequencies of the input signal.

The operational characteristics of the proposed model require us to consider the optimisation of the step which digitize the analysed input signal. Assuming that the measurement of this signal has been made with appropriate parameters it is necessary to select a step that allows for possibly the most precise signal mapping in each case. This will make the created hysteresis loop equally precise. Taking into

account the character of the input signal the number of steps will be directly proportional to its growing frequency.

If the conditions concerning the input signal and resulting from the assumption of hysteresis are met then the model will generate hysteresis loop dependent on/independent from the frequency of the input signal and always dependent on the value of the input function.

Having defined the magnetostatic hysteresis, which emerges as a result of magnetoelastic effects and whose nature and diagnostic usability have been also verified experimentally, one can move to testing the possibilities offered by the proposed model of hysteresis in terms of the description of its characteristic behavior. In this case one ought to present the characteristic behavior of hysteresis loop during stretching of the sample in the range of elastic deformations. As has been proven experimentally, in the case of this range there occurs change of the area of the loop's surface as the amplitude of the exciting input signal changes. The simulation of such behavior, while using the verified model, is presented in Fig. 6. The loops emerge in the plane of the input (the run of the exciting force) and the output (changes of a sample's own magnetic field).

The equation was made for two different load amplitudes, respective equal to 4 and 8 kN. Application of the model for this specific case required a scale factor to be introduced. This factor defined the extent of the change of the areas of a loop's surface proportionately to the amplitude of the input.

We have presented above the quality profile of the model's functioning. Quantity description will depend on internal parameters of model. It is also possible to introduce mathematical rules into this original model and they will additionally provide corresponding values for each of hysteresis loop's arguments.

The developed model has the capability of performing the modelling of the hysteresis' characteristics. These characteristics may be used for interpreting magneto-mechanical effects. The model functioning in the form presented herein is ready for the introduction and verification of physical correlations, which will lead to a better representation of real behaviors.

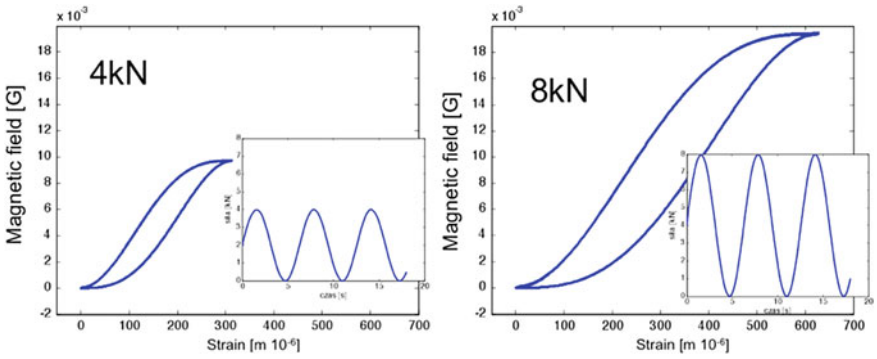


Fig. 6 Simulation of the hysteresis loop for various amplitudes of the force in the range of elastic deformations

6 Conclusion

Basing solely on the influence of the weak Earth's magnetic field, it was established that it is possible to represent stresses by means of the measurement of magnetic field's changes generated around the object made of materials of ferromagnetic properties. On the basis of the information gathered and the analysis directed towards the application of the hysteresis diagnostic model in diagnosing the states of stress, it was found that it is possible to transform a signal to a given shape of hysteresis that has the potential of identifying signal's parameters changes.

The developed model turned out to be capable of reflecting the physical nature of observed phenomena, by demonstrating the properties of hysteresis which can be used for interpreting magneto-mechanical effects. The model, in the presented, operating form, is ready for inputting and verifying the physical relations, which will result in better representation of actual behavior. Generally a concept has been presented here of using hysteresis as the base for a diagnostic model. The results point to high potential of such an approach to diagnosis of objects which are characterized by dynamic response.

The idea of using hysteresis in diagnostic and prognostic procedures of the material's (construction) stress state analysis seems to be an interesting proposal for SHM systems. An exemplary procedure of a construction's diagnosis could consist in using magnetic measurement system for finding the critical object's nodes, and then constant monitoring of borderline stresses while taking into account the range of permanent and elastic deformation, which would allow for early detection of threats. A series of magneto-mechanical and other phenomena occurring during construction/operation of a machine as well as the actual magnetization conditions of the object which was examined in an external magnetic field (Earth's field or the field derived from adjacent elements) will be taken into consideration in SHM data analysis. This approach enforces collaboration with a correspondent mathematical-physical model. The proposed model could meet this challenge.

The issues discussed in the present chapter give hope for development of a passive magnetic method of diagnosis which exploits the change of an object's own magnetic field to provide new opportunities for detecting the stress which could damage construction objects made of ferromagnetic material.

Acknowledgments This work was supported by The National Centre of Research and Development (Poland) within grant no. PBS1/B4/6/2012. This work has been also supported by the European Union in the framework of European Social Fund through the Warsaw University of Technology Development Programme, realized by Center for Advanced Studies.

References

1. Lindgren M, Lepistö T (2003) Relation between residual stress and Barkhausen noise in a duplex steel. NDT E Int 36(5):279–288
2. Bao S, Gong SF (2012) Magnetomechanical behaviour for assessment of fatigue process in ferromagnetic steel. J Appl Phys 112(11):113902
3. Jiles DC (1995) Theory of the magnetomechanical effect. J Phys D: Appl Phys 28:1537–1546
4. Salach J, Bienkowski A, Szewczyk R (2007) Magnetoelastic, ring-shaped torque sensors with the uniform stress distribution. J. Automat Mobile Robot Intell Syst 1:66
5. Xingliang J, Xingchao J, Guoyong D (2009) Experiment on relationship between the magnetic gradient of low-carbon steel and its stress. J Magn Magn Mater 321(21):3600–3606
6. Sz Gontarz, Radkowski S (2012) Impact of various factors on relationships between stress and eigen magnetic field in a steel specimen. IEEE Trans Magn Magn 48(3):1143–1154
7. Florianowicz M, Bohdal L (2012) Modeling of the refrigerants condensation in the superheated vapor area. Ann Set Environ Prot 14:393–406
8. Ikhouane F, Gomis-Bellmunt O (2008) A limit cycle approach for the parametric identification of hysteretic systems. Syst Control Lett 57:663–669
9. Ahrens J, Tan X, Khalil HK (2007) Multirate sampled-data output feedback control of smart material actuated systems. In: Proceedings of the American control conference, New York
10. Smith RC, Dapino MJ, Braun TR, Mortensen AP (2006) Homogenized energy framework for ferromagnetic hysteresis. IEEE Trans Magn 42(7):1747–1769

Online Monitoring of Steel Constructions Using Passive Methods

Szymon Gontarz, Jędrzej Mączak and Przemysław Szulim

Abstract In the chapter the possibilities of development of distributed diagnostic system capable of online structural health monitoring of civil engineering structures are discussed. Instead of commonly used methods of determining the technical state of the construction that are usually focused on searching for cracks, material heterogeneities and assessing concrete or steel degradation, the methods proposed for the system are based on the comparative strain gauge, acceleration and passive magnetic measurements. All this measurement methods are used for stress assessment in critical fragments that are vital for stability and durability of the structure. Evolution of defects in the construction causes measurable changes of stress distribution in critical joints of the construction. Additionally, materials that could cause threat of the catastrophic accident caused by fatigue wear, exceeding stress limits or emerging of plastic deformations have magnetic properties that could affect the local magnetic field. Stress limits could be then constantly supervised taking into account the range of permanent and elastic deformation, which would allow for early detection of threats. SHM data analysis is taking into consideration a series of magneto-mechanical and other phenomena occurring during the construction/machine's operation and also actual magnetization conditions of the object examined in exterior magnetic field (Earth's or derived from adjacent elements).

S. Gontarz (✉) · J. Mączak · P. Szulim
Integrated Laboratory of the Mechatronics System of Vehicles and Construction Machinery, Warsaw University of Technology, Warsaw, Poland
e-mail: sgontarz@simr.pw.edu.pl

J. Mączak
e-mail: jma@simr.pw.edu.pl

P. Szulim
e-mail: p.szulim@mechatronika.net.pl

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_54

1 Introduction

Distributed diagnostic systems are widely used in machine diagnostic for monitoring condition of critical machines e.g. power units, fans, etc. allowing for on-line monitoring and performing exploitation decisions according to the current state of the monitored objects [1–3]. The main advantage of this approach is the possibility of simultaneous monitoring technical state of many objects distributed on a large area from one place thus limiting costs and manpower. Especially it affects cases when there are long distances between diagnosed objects and a diagnostic technician. This distance is thus limited only by the network availability and its performance. This concept could be adopted for on-line monitoring of civil infrastructure objects (Fig. 1).

Usually, online diagnostic systems monitoring structural health of civil engineering structures are focused on stress assessment in the construction. For this purpose various methods of technical state assessment could be used based on the comparative dynamic, tensometric (strain gauges), magnetic and optic fibres (FBG) measurements. All this measurement methods allowed for stress assessment in critical fragments that are vital for stability and durability of the structure.

Exemplary topology of the distributed SHM system was shown on Fig. 2. The system shown is built around cRIO (National Instruments) controllers that are monitoring critical parts of the construction. Controllers are used for signal acquisition from sensors (magnetic, strain gauge and optic FBG sensors), processing data and evaluating stress in the elements of the construction.

Evolution of defects in the construction, specifically exceeding the yield point of the material, causes measurable changes in the dynamic properties along with evolution of stress distribution in critical construction joints. Additionally, materials that could cause threat of the catastrophic accident caused by fatigue wear, exceeding stress limits or emerging of plastic deformations have magnetic properties that could affect the local magnetic field [4–6].

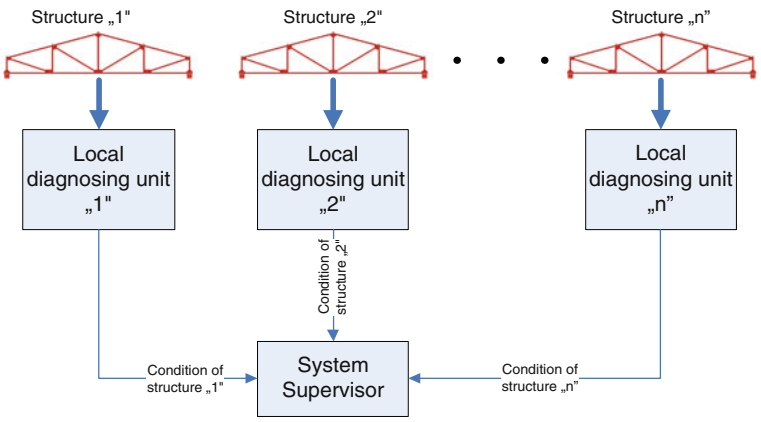


Fig. 1 General layout of the distributed diagnostic system [3]

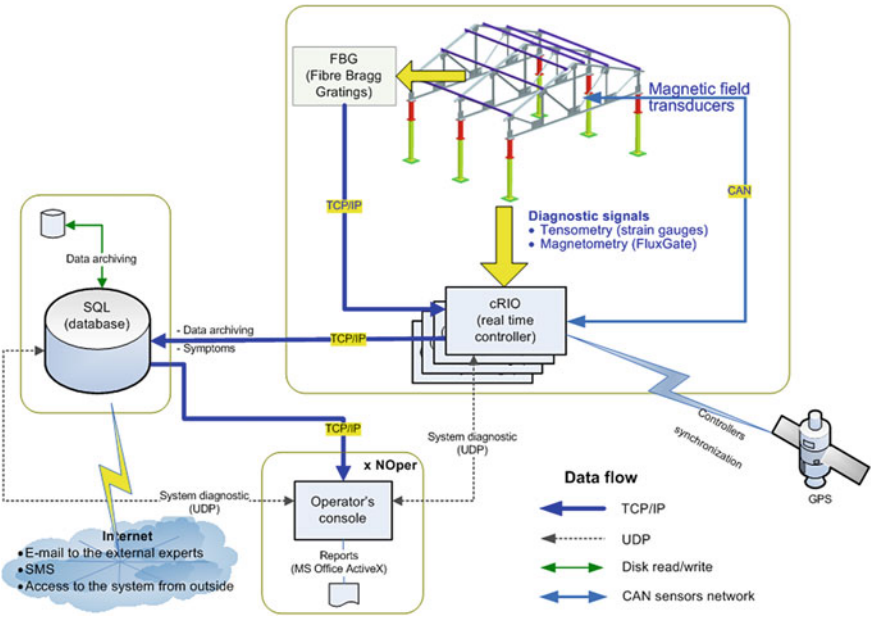


Fig. 2 Topology of the distributed diagnostic system for monitoring the construction

2 Description of the Passive Magnetic Method

The idea of using passive magnetic method in diagnostic and prognostic procedures of the assessment of the material's (construction) stress seems to be an interesting proposal for SHM systems [3]. Materials that have a high risk of damage due to material fatigue, exceeding maximum stress or plastic deformations have magnetic properties, which allowed developing a group of magnetic methods in technical diagnostics. Currently, a group of passive diagnostic methods is developing dynamically in parallel to active diagnostic methods [6, 7]. The group of passive diagnostic methods has all the advantages of the group of the active diagnostic methods, and additionally it does not require the application of artificial source of magnetic field that needs sophisticated and expensive equipment.

An exemplary procedure of a construction's diagnosis consists in using magnetic measurement system for finding critical object's nodes, and then constant monitoring of stress limits that take into account the limits of permanent and elastic deformation, which would allow for early detection of threats. SHM data analysis shall take into consideration a series of magneto-mechanical and other phenomena (e.g. the Villary, Matteucci and Naganka-Honda effects) occurring during the construction/machine's operation and also actual magnetization conditions of the object examined in exterior magnetic field (Earth's or derived from adjacent elements). This forces collaboration with a corresponding mathematical-physical model.

In contrary to the strain gauge sensors that are measuring stress in the particular point of the construction magnetic sensors are measuring changes in the magnetic field of the construction node from a certain distance. Therefore the obtained results correspond to the total stress in the node or the part of the construction. A comparative test of the results obtained for stress assessment in the constructions, with the use of different passive methods (optical, strain gauge and magnetic), could be find in [8].

3 Sensors for Magnetic Field and Acceleration Measurements of the Construction

Sensors typically used for the magnetic field measurements (e.g. Applied Physics Systems APS-536 fluxgate transducers) are rather expensive so they could be used sporadically in the online diagnostic system or for other magnetic sensor calibration. APS-536 is a small magnetic sensor that allows for measuring all three perpendicular components of the magnetic field vector simultaneously. It could be used for measuring magnetic field in the range of 3×10^{-7} to 1 G (Gauss).

On the market many general purpose, low budget, magnetic OEM single chip transducers could be found. They could be used for developing magnetic field sensors dedicated for online diagnostic system of the construction. Below short characteristics of the selected transducers is presented.

MAG3110 transducer is an integrated, digital, three axial 16 bit transducer produced by Freescale [10]. It has a single measuring range of ±1000 μT with resolution of 0.03 μT and eight sampling rates ranging from 0.63 to 80 Hz. It should also be noted that the signal-to-noise ratio of the transducer is very low. Integrated electronics takes care of sampling the analogue signal at the programmed frequency. Measurement data can be accessed via the I2C bus. Noteworthy is the fact that of the power separation of the analogue and digital part of the sensor. With this separation the level of noise in the measured signal could be greatly reduced. Additional integrated temperature sensor provides data for temperature correction parameters, like sensitivity, or the zero level, which are more or less susceptible to temperature changes.

Another notable transducer is an analogue transducer codenamed AAH002 produced by NVE Corporation Company that is using magneto-resistance giga-effect [11]. It has a relatively high sensitivity and is working in the unipolar configuration. This is undoubtedly a shortcoming and also a characteristic feature of this type of transducers. The sensor can correctly measure only negative (from negative values to 0) or positive (from 0 to positive values) magnetic field, the feature that is impractical in automatic measuring system. The transducer contains a bridge composed of measuring elements (Fig. 3). Bridge output voltage varies in proportion to the value of the applied magnetic field. The fact that it is an analogue transducer allows application of a set of external antialiasing amplifiers and filters.

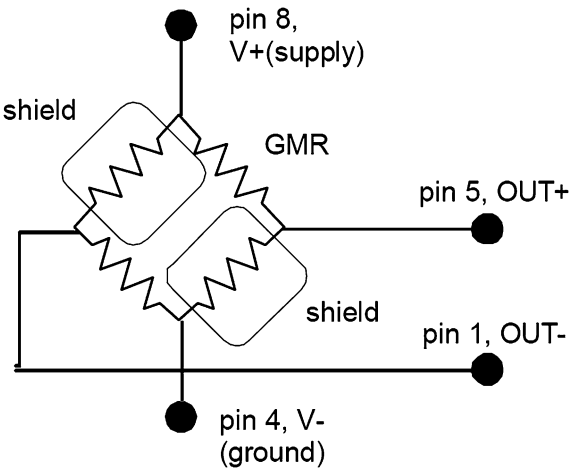


Fig. 3 Electric layout of the AAH002 transducer [11]

Therefore it should be possible to obtain a much higher sensitivity and a better cut off from the noise measurements. What's more, in theory it becomes possible to build lower-cost differential magnetic field sensor.

Another interesting option are the RM3000 and RM2000 transducers produced by the PNI Sensor Corporation company [12]. These relatively cheap sensors are based on the hysteresis effect occurring in the ferromagnetic core. The company provides a single analogue sensors and ASICs circuits that integrated analogue and digital parts. For these sensors worth considering is a selection of a ready-integrated solution, because analogue sensors require a fairly complex control system in order to obtain the measurement. The measuring range is similar to the previously mentioned transducers. An important advantage of the transducer is relatively high resistance to temperature changes. This is a very important feature, since in many applications, sensors have to operate in an environment of changing parameters such as temperature, humidity sensor, etc. Insensitivity to changes in these parameters helps to reduce measurement errors.

In order to test the possibility of using the passive construction diagnostics based on the measurement of the magnetic field sensors a new sensor capable of measuring magnetic field and acceleration of the construction was developed (Fig. 4). The sensor consists of several key components. The main measuring element is an integrated three-axis magnetic field transducer LSM303 produced by STMicroelectronic. This transducer also provided three-axis acceleration sensor. Basic technical parameters of the transducer could be found in [13].

The signal from the bridges is subjected to pre-amplification and processed with 12-bit A/D converter. This signal is available to host computer/microcontroller via I2C bus. This bus is a typical bus for communication between two devices operating in close proximity. It is not suitable for the transmission of data over a long distance. Therefore the developed sensor controller, in addition to communication with the LMS303 sensor also performs the function of sending measurement data

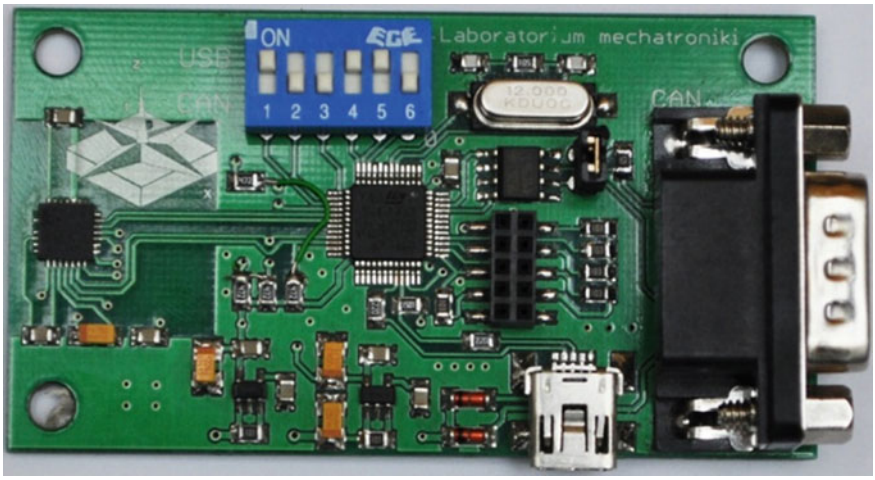


Fig. 4 Sensor for magnetic and acceleration measurements based on LSM303

over the CAN bus. The magnetic sensor uses 7 sensitivity levels. In a variant of maximum sensitivity the measuring range is ±1.3 g_n and the measurement resolution is of 1 mg_n. The maximum range of the sensor is ±8.1 g_n. An important issue in this type of measurement is the frequency of measurement. Developed sensor allows for 7 different sample frequencies ranging from 0.75 to 70 Hz.

During the tests it was discovered that the antialiasing filter of LMS303 could cause problems especially with filtering disturbances caused by the power systems.

An important feature of LSM303 transducer is that it has been calibrated at the factory. This calibration refers to the sensitivity axis and the influence of temperature on the sensitivity and the offset of the sensor. Offset in this case is understood as a certain DC component that appears in the measured signal. It must be remembered that the factory calibration relates only to LSM303. The sensor is built into the device and, due to the presence of ferromagnetic materials, requires re-calibration. It should also be noted that the sensor axes need not be perpendicular to each other and the axis position errors can be as high as several degrees. These parameters are also a subject to calibration.

An interesting feature of the constructed device is the ability to measure acceleration. The built-in three-axis accelerometer sensor LSM303 was made in MEMS technology. It has sampling frequency of 1 kHz, three measurement ranges: ±−2, ±4 and ±8 g_n (standard acceleration due to gravity), and quite good temperature parameters. Fairly narrow range causes that it is impossible to use the sensor for measuring vibration of the structure. An interesting application is, however, to use the sensor for measuring the deformation of the structure. Sensors made with this technology allow measuring of earth gravity. This measurement is similar to other types of sensors and is performed indirectly by measuring the force acting on a reference mass due to the presence of acceleration. Because the sensor

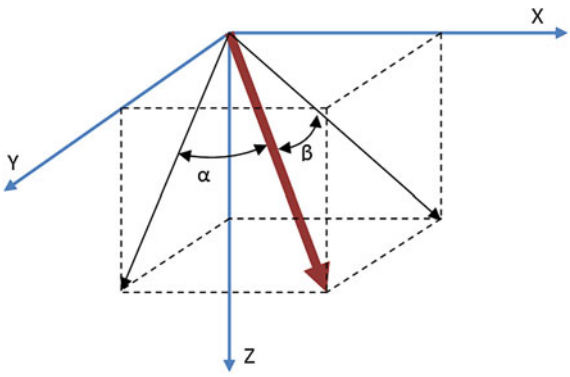


Fig. 5 Coordinate system for measuring rotation angles

measures the value of the DC component of acceleration, assuming that the sensor is fixed, you can measure the acceleration of gravity. Since its value may be treated as a constant, the measurement of the acceleration vector can be used to determine the angle of rotation of the sensor. Transducer arranged in the measuring device, attached to a test structure can easily be used to detect deformations (for example by buckling) of its elements. The following figure (Fig. 5) is an example of the sensor coordinate system with the selected acceleration vector and two angles describing the deviation vector and thus a rotation of the sensor line, perpendicular to the ground.

It is worth noting that the resolution of the ADC 12 bits and sensitivity of 2 g_n, obtained acceleration measurement resolution is about 1 mg_n. With such a resolution, for small values of the angles α and β, the obtained angle measurement resolution is of 0.05°. The usual problem is the measurement noise, so in order to achieve a more accurate measurement it is suggested to use the analogue sensors and refined analogue measurement path.

4 Test Results

In this section the results obtained during the experiment of truss loading will be presented. A force was applied to the steel truss structure [9] and the strain in the selected joints of the construction as well as its deflection using LVDT sensor was measured. The truss was equipped with a set of 8 low cost sensors as of Fig. 4 measuring changes of the magnetic field and accelerations in selected points of the construction (Fig. 6). During the experiment the load was gradually increased several times using two hydraulic jacks from 0 to 35 kN on each jack. In the end of the experiment the structure was damaged by buckling after applying load of 35 kN (on each jack).

Figure 7 presents a graph of changes in the angle β in the three measurement nodes during the experiment. On the top graph the truss loading is presented. Simultaneously with magnetic field and acceleration the strain in the selected points

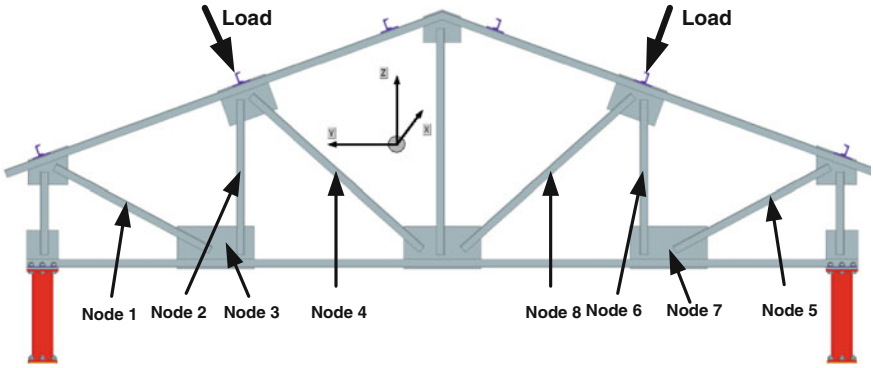


Fig. 6 Location of sensors

of the construction was measured. Figure 7 shows that in the first part of the experiment when the force varied in the range from 10 to 30 kN, the rotation angle recorded in the nodes practically did not change.

The buckling of the truss was first visible on Node 2 after 4,000 s of the start of the experiment (force 45 kN) as a sudden change of the indications. During the next loading cycles this effect was visible even better as the truss was already weakened.

Figure 8 shows a graph of the registration of angles α and β in the Node 6. It shows that the angle α has not changed substantially during the experiment. Due to the nature of the buckling angle α changes were small because omitted from the presentation of the results.

Figures 9 and 10 shows the changes of the modulus of the magnetic field vector in Nodes 1–3 and Nodes 5–7. Clearly visible are changes of magnitude of the magnetic field in the moment of buckling

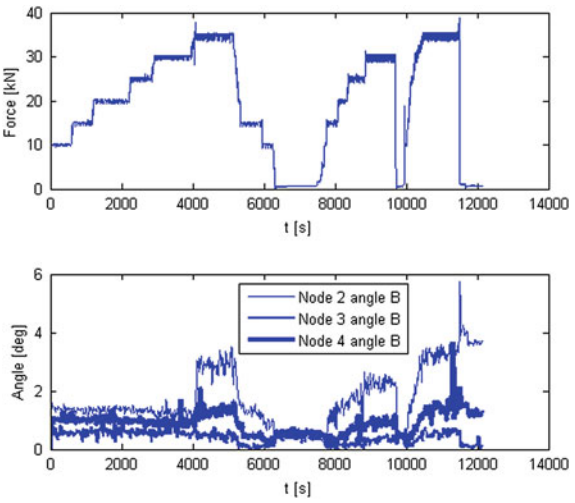


Fig. 7 Influence of loading on changes of angle β

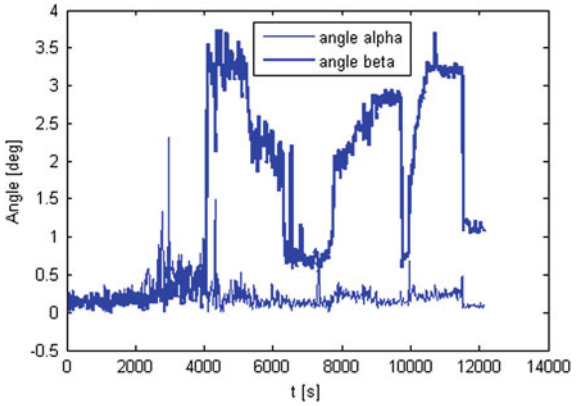


Fig. 8 Change of angles α and β in the Node 6

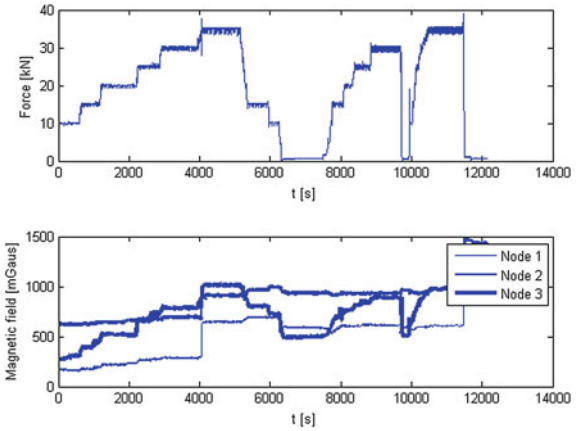


Fig. 9 Changes of the modulus of the vector of the magnetic field for Nodes 1, 2 and 3

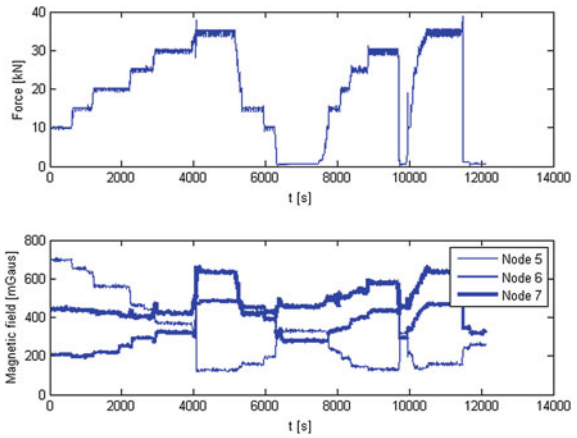


Fig. 10 Changes of the modulus of the vector of the magnetic field for Nodes 5, 6 and 7

In Figs. 9 and 10 can be seen a clear correlation of the measurement signal derived from a magnetic field sensor with a value set by the actuator force. Closer analysis shows an almost linear relationship until the point of buckling, which occurred about 4,000 s of the experiment. Then there was a sudden change in the magnetic field recorded in all measurement nodes. After buckling and the removal of the loading from the structure, the value of the magnetic field in measurement nodes did not return to its original state but remained on some specific levels. It is closely related to the strain that resulted from buckling. After buckling the biggest differences of the magnetic field were observed in measurement nodes located near the points where the plasticity of the material was greatest. Because the strut structure was permanently deformed, another load cycle caused the different behaviour of the structure and thus different magnetic response to the changing load. In Fig. 10 the course of changes of the magnetic field recorded in the node 7 varied slightly with increasing load in a first cycle (0–4,000 s). The second load cycle (from 6,000 to 10,000 s) shows an increase of the modulus of the magnetic field with similar loading force.

The geographic location in which the experiment was carried out an average absolute value of the vector Earth's magnetic field is about 440 mG. Figures 9 and 10 shows that the value of the magnetic field measured on the structure at the beginning of the experiment were very different from each other and from the Earth's field. It is related to the magnetization of a typical structure for this kind of heat-treated steel with additional mechanical treatment. Figure 9 shows that at some nodes the magnetic field has reached a value of 1,500 mG while in the other nodes it has much lower value. In some nodes the loading applied to the structure caused an increase or decrease of the magnetic field which was associated with a number of phenomena taking place in the steel structure eventually causing not only increased of the magnetic field vector associated with the effects of magneto elastic but also its rotation. The resultant magnetic field vector which is the vector sum of the fields coming from the construction and the Earth's field can both decrease and increase due to increase of the magnetic field of the construction. This requires taking into account of the many initial conditions.

5 Conclusion

Adaptation of the distributed diagnostic systems technology widely used for diagnosing mechanical systems for the purpose of monitoring critical elements of the infrastructure objects is very promising as it could improve the safety of that objects and will lower the probability of catastrophic events.

The idea of using passive magnetic method in diagnostic and prognostic procedures of the material's (construction) stress state analysis seems to be an interesting proposal for SHM systems. In the chapter some preliminary results of using low-cost magnetic field and acceleration transducers were shown. An exemplary procedure of a construction's diagnosis could consist in using magnetic measurement system for

finding critical object's nodes, and then constant monitoring of stress limits that take into account the range of permanent and elastic deformation, which would allow for early detection of threats. SHM data analysis shall take into consideration a series of magneto-mechanical and other phenomena occurring during the construction/machine's operation and also actual magnetization conditions of the object examined in exterior magnetic field (Earth's or derived from adjacent elements).

Acknowledgments This work was supported by The National Centre of Research and Development (Poland) within grant no. PBS1/B4/6/2012. This work has been also supported by the European Union in the framework of European Social Fund through the Warsaw University of Technology Development Programme, realized by Center for Advanced Studies.

References

1. Maczak J (2007) Structure of distributed diagnostic systems as a function of particular diagnostic task. In: 20th International congress on condition monitoring and diagnostic engineering management, Faro, Portugal, pp 171–178
2. Maczak J (2013) The concept of the distributed diagnostic system for structural health monitoring of critical elements of infrastructure objects. In: Amadi-Echendu JE et al (ed) Asset condition, information systems and decision models. Springer London
3. Gałęzia A, Gontarz S, Jasiński M, Maczak J, Radkowski S, Seńko J (2012) Distributed system for monitoring of the large scale infrastructure structures based on analysis of changes of its static and dynamic properties. Key Eng Mater 518:106–118
4. Gontarz S., Radkowski S (2012). Impact of various factors on relationships between stress and Eigen magnetic field in a steel specimen. IEEE Trans Magn 48(3):1143–1154. doi: 10.1109/TMAG.2011.2170845, 2012
5. Bao S, Gong SF (2012) Magnetomechanical behavior for assessment of fatigue process in ferromagnetic steel. J Appl Phys 112(11):113902
6. Lindgren M, Lepistö T (2003) Relation between residual stress and Barkhausen noise in a duplex steel. NDT E Int 36(5):279–288
7. Dubov A (2004) Principal features of metal magnetic memory method and inspection tools as compared to known magnetic NDT methods. In: Proceedings of the 16th annual world conference on non-destructive testing, Montreal, Canada
8. Radkowski S, Gontarz S, Maczak J, Kujawińska M, Dymny G, Malowany K (2011) Experimental comparative testing of steel structures by strain gauges, digital Image Correlation and magnetic field methods. In: Proceedings of the eighth international conference on condition monitoring and machinery failure prevention technologies, Cardiff
9. Maczak J (2012) A structural health monitoring system based on an analysis of changes in the static, dynamic and magnetic properties of the structure. In: Topping BHV (ed) Proceedings of the eleventh international conference on computational structures technology, Civil-Comp Press, Stirlingshire, UK, Paper 86. doi:10.4203/ccp.99.86
10. Freescale. http://www.silica.com/fileadmin/02_Products/Productdetails/Freescale/SILICA-Freescale-MAG3110-ds.pdf. Accessed 13 July 2013
11. NVE Corporation. http://www.rhopointcomponents.com/images/aaabsensors.pdf. Accessed 23 Aug 2009
12. PNI Sensor Corporation. http://www.pnicorp.com/products/rm3000-rm2000. Accessed 13 July 2013
13. STMicroelectronics. http://www.st.com/web/catalog/sense_power/FM89/SC1449/PF251902. Accessed 13 July 13 2013

Experimental Research on Misfire Diagnosis Using the Instantaneous Angular Speed Signal for Diesel Engine

Yu-hai He, Jian-guo Yang, Cheng'en Li and Fu-song Duan

Abstract Built a test bench for WP10.240 high-speed diesel engines, the normal and misfire faults were simulated, the Top Dead Centers (TDC) signal and Instantaneous Angular Speed (IAS) signal were measured; The IAS signals of the diesel engine were processed by the period average, tooth average and smooth handling methods, the IAS signals were analyzed in both conditions of normal and single cylinder misfire, the feature parameters were abstracted, and the fault diagnosis criteria using the IAS signal for misfire were obtained.

Keywords Misfire diagnosis · Instantaneous speed · Diesel engine

1 Introduction

As high-speed diesel engines are widely used as power plants in engineering machines, generators and trucks, the high requirements of the safety and reliability should be satisfied. Due to the complicated structure, various and serious working conditions, and the long-term continuous operating of high-speed diesel engine, it is easy to cause a economic loss and to put staff's safety in danger once fault occurs. With the fast development of the computer science and the information technology, the diesel engines develop in the direction of automation and integration, and the monitoring and diagnosis methods is improving better. Timely potential failure

Y. He (✉) · J. Yang · C. Li · F. Duan
School of Energy and Power Engineering, Wuhan University of Technology, Wuhan 430063, China
e-mail: hyh@whut.edu.cn

J. Yang
e-mail: jgyang@whut.edu.cn

Y. He · J. Yang
Key Laboratory of Marine Power Engineering and Technology under Minister of Communication, Wuhan 430063, China

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_55

information of the diesel engine is monitored and the measures are taken. So the safety and reliability of the diesel engine is improved and economy is assured.

Instantaneous speed is an important parameter to evaluate the power performance, torque characteristics and reliability performance for the diesel engine. During operation of the diesel engine, the dynamic performance of each cylinder is basically the same. Although there are differences between instantaneous speed fluctuations, this difference is always in a small range and presents some regularity. However, when a cylinder fails, the engine power of each cylinder balance is destroyed [1] and the instantaneous speed signal will be distorted, which can determine the operating condition of the engine cylinder. Measurement of the instantaneous speed signal is convenient, the sensor installation is simple, monitoring the instantaneous speed does not affect the normal operation, and it can identify the faulty cylinder [2, 3], especially for online monitoring and diagnosis.

In the thesis, WP10-240 diesel engine selected as the research object, experimental research on misfire diagnosis using the instantaneous angular speed signal for diesel engine was carried out, the fault criteria of the instantaneous speed for misfire is obtained.

2 The Measuring Principle of the Instantaneous Speed

The 360-degree crank angle (crank rotation one week) is divided into equal space interval, which can be measured by the sensor to get the time-domain waveform of every space interval. Calculating the elapsed time you can get the instantaneous speed. Currently, the instantaneous speed measurement methods are mainly magneto-electric method and photoelectric method [4]. In photoelectric method, for example, a reflective tape is generally arranged on the shaft or flywheel, the photoelectric encoder will convert optical signals into electrical signals, which is usually a rectangular pulse. During operation of the engine, the photoelectric sensor output pulse signals, and each pulse signal corresponds to one of the teeth on the flywheel space interval. Shown in Fig. 1, the instantaneous speed corresponding to current space interval is V_i.

$$v_i = \frac{60 \cdot f_s}{[z \cdot (p_{i+2} - p_i)]} \tag{1}$$

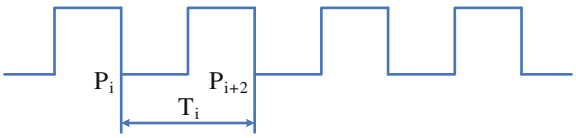


Fig. 1 The measuring principle of the instantaneous speed

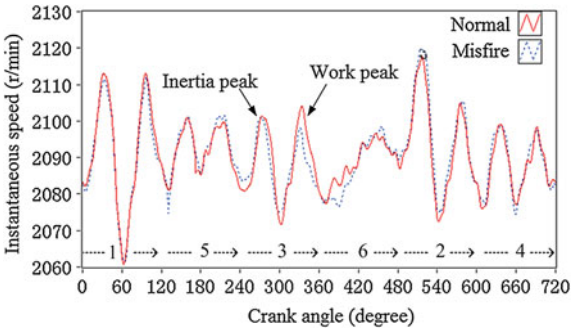


Fig. 2 The speed difference of NO. 3# cylinder under normal operating conditions

Where fs is the sampling rate, z is the number of pulses per revolution, P_i is the intersection between the original instantaneous speed and the zero line.

In Fig. 1, P_i and P_{i+2} represent intersection between the original instantaneous speed and the zero line. T_i is time interval of a period.

3 Fault Diagnosis Mechanism and Fault Feature Extraction

Through analysing the instantaneous speed waveforms of the four-stroke engine, we can find there are two instantaneous speed peaks corresponding in each cylinder working area. The first peak corresponds to the instantaneous speed's inertia peak, the second peak corresponds to the instantaneous speed's work peak [4]. If a cylinder's work ability decreased because of the fault, the second instantaneous speed peak is accordingly reduced, which can determine the type and extent of the fault resulted in the power drop, and according to TDC signal, we can locate the fault cylinder.

Instantaneous speed fluctuations is the result of gas pressure, reciprocating inertia force and so on [5]. Analysed the waveform of instantaneous speed, the speed difference is to be taken as a fault characteristic parameters. The difference between the instantaneous speed's work peak and the inertia peak, is defined as the speed difference (Ins_Dif_Cyl.i) corresponding to i# cylinder. The speed difference is calculated as follow formula, shown in Fig. 2.

$$\text{Ins\_Dif\_Cyl.i} = \text{Average.i2} - \text{Average.i1}$$

Where: Average.i2 is the mean value of work peak around the NO. i# cylinder within a certain crank angle; Average.i1 is the mean value of inertia peak around the NO. i# cylinder within a certain crank angle.

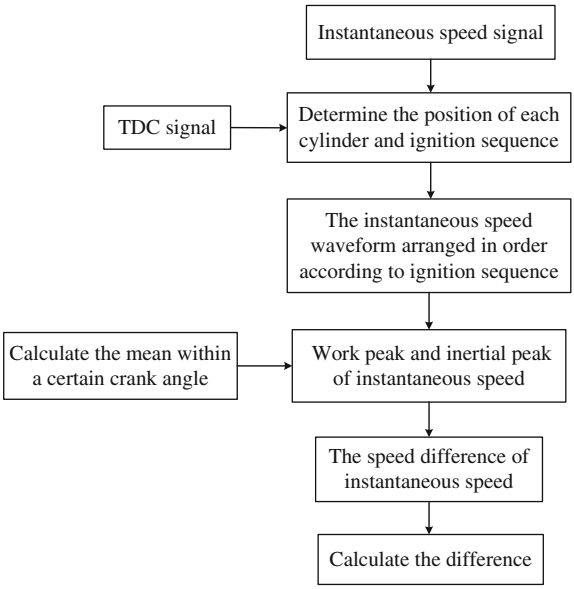


Fig. 3 Instantaneous speed signal processing

4 Data Processing

Since the structure and working characteristics of diesel engines, there will be fluctuations in the operating state as the speed, load and other conditions change,the instantaneous speed signal will fluctuate in the face of the same conditions. In order to eliminate the random measurement error [6], and improve the accuracy of calculation, the instantaneous speed curve would be more smooth and continuous manner, it's essential for instantaneous speed signal to signal process, such as tooth average [7], multi-cycle average [8], and digital filtering [9].

Instantaneous speed signal processing was shown in Fig. 3. First, to measure the instantaneous speed signal under normal operating conditions, and then measure the instantaneous speed signal under fault condition (simulation). For comparative analysis, it is necessary for average speed under fault condition to process, and so its value equal to the average speed under the normal operating conditions, i.e., the instantaneous speed signal waveform under fault condition will pan up and down.

5 Experimental Research on Misfire Diagnosis

5.1 The Object of Test

Based on the current laboratory conditions of School of Energy & Power Engineering, Wuhan University of Technology, the WP10-240 diesel engine was chosen. The details of the diesel engine are listed in Table 1.

Table 1 WP10-240 diesel engine parameters

Cylinder number	6	Firing order	1–5–3–6–2–4
Cylinder bore	126 mm	Piston stroke	130 mm
Rated power	175 kW	Rated speed	2,200 r/min
Maximum torque	1,000 N·m	Compression Ratio	17:1

Table 2 artificial faults

Test no.	Simulation of fault type
1	Normal
2	Cly. 1 misfire
3	Cly. 2 misfire
4	Cly. 3 misfire
5	Cly. 4 misfire
6	Cly. 5 misfire
7	Cly. 6 misfire

5.2 Artificial Faults of the Diesel Engines

The artificial faults created in Cyl.6 of WP10-240 diesel engine for testing are specified in Table 2.

5.3 Working Conditions

The WP10-240 diesel engine worked at 2,100 r/min with no load. The diesel engines will be tested at above case respectively.

5.4 The Location of Measuring Points

In order to acquire the basic parameters related to the condition of diesel engine, many signals have been measured, such as TDC, crank angle etc. and the speed sensor is selected the optical encoder(1,024 pules per cycle, DC5 V). According to the acquisition system setup, the locations of the measuring points for WP10-240 diesel engine were shown as Fig. 4 and Fig. 5.

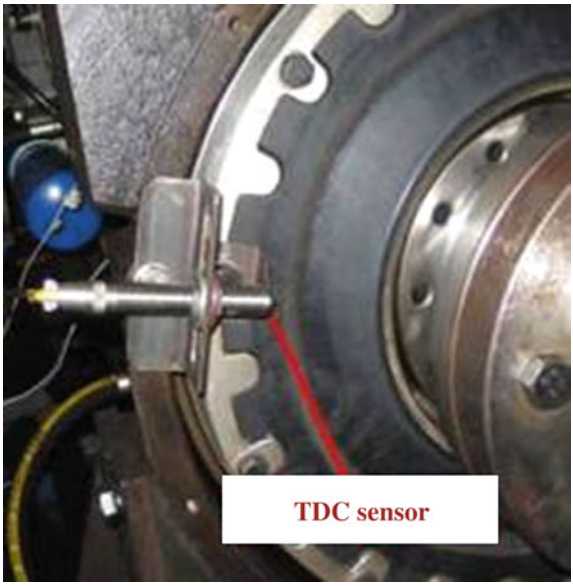


Fig. 4 The location of TDC measuring points

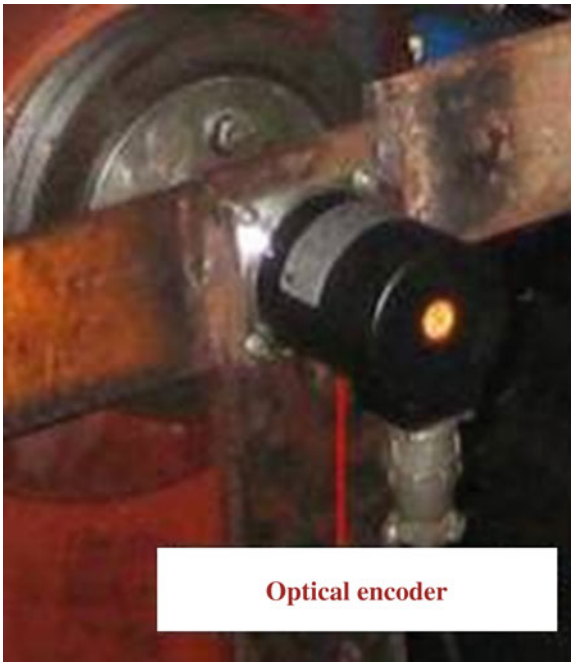


Fig. 5 The location of optical encoder measuring point

6 Test Results

The single-cylinder misfire fault simulation tests have been done in WP10.240 diesel engine, the instantaneous speed signals have been acquired, and compared with the normal working conditions, the test results are shown in Figs. 6, 7, 8, 9, 10 and 11.

The difference of speed difference between single-cylinder misfire condition and normal condition is shown in the following Table 3.

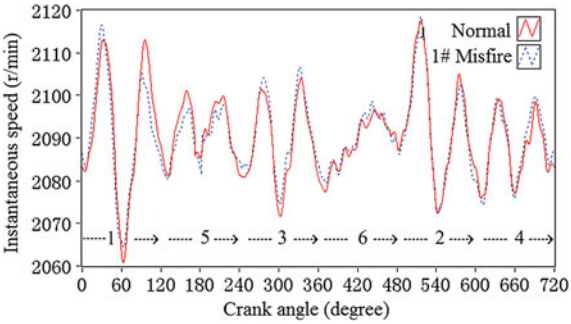


Fig. 6 The instantaneous speed comparison between normal condition and 1# misfire fault

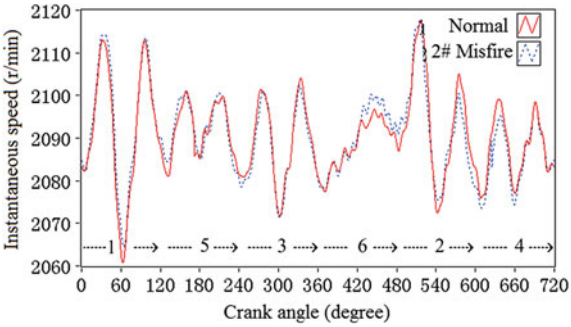


Fig. 7 The instantaneous speed comparison between normal condition and 2# misfire fault

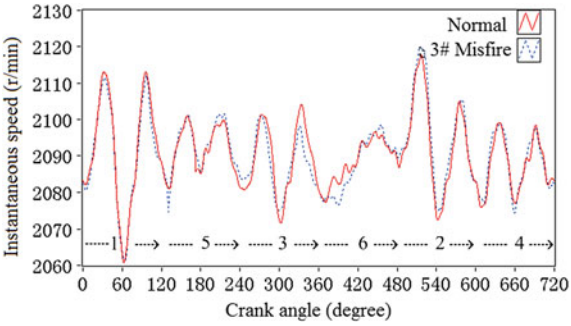


Fig. 8 The instantaneous speed comparison between normal condition and 3# misfire fault

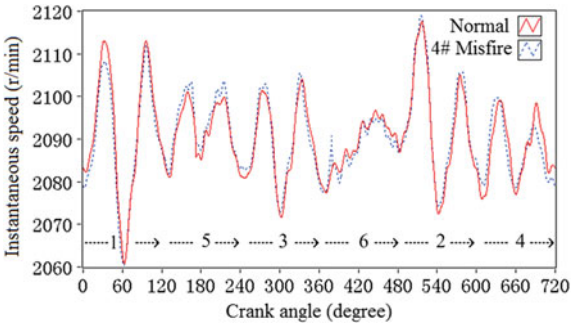


Fig. 9 The instantaneous speed comparison between normal condition and 4# misfire fault

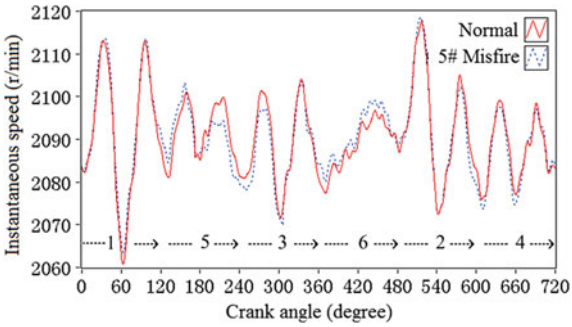


Fig. 10 The instantaneous speed comparison between normal condition and 5# misfire fault

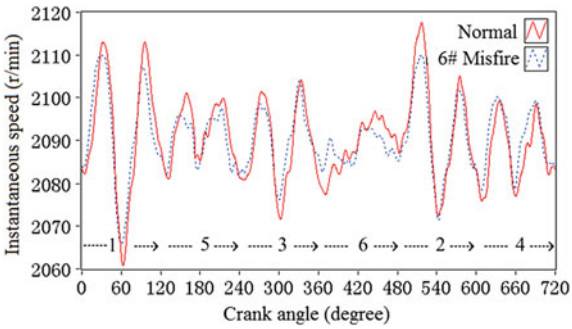


Fig. 11 The instantaneous speed comparison between normal condition and 6# misfire fault

From the above figures of the instantaneous speed waveform and table of speed difference, it shows when a cylinder misfired, compared with the normal working conditions, its work peak of instantaneous speed would reduce and inertial peak of instantaneous speed would increase, while the other cylinders did not change significantly. In this experiment, the cylinder which speed difference is absolute minimum may have misfired. Therefore, the speed difference of instantaneous speed can be used to monitor and diagnose whether a cylinder misfired, which can be judged using the following criteria:

Table 3 The difference of speed difference between single-cylinder misfire and normal condition

Data analysis Condition	Ins_Dif_Cyl. 1	Ins_Dif_Cyl. 5	Ins_Dif_Cyl. 3	Ins_Dif_Cyl. 6	Ins_Dif_Cyl. 2	Ins_Dif_Cyl. 4
Cyl.1 misfire	−4.77	1.46	1.17	0.21	−0.69	3.47
Cyl.2 misfire	0.49	−0.32	0.59	3.92	−3.52	4.36
Cyl.3 misfire	1.15	1.46	−3.57	6.64	−0.95	−0.65
Cyl.4 misfire	4.42	−0.16	−1.29	−0.35	1.61	−3.99
Cyl.5 misfire	1.1	−6.01	3.38	1.23	−2.29	1.1
Cyl.6 misfire	−3.85	0.64	3.24	−7.25	3.83	2.12

(1) The speed difference is absolute minimum;
(2) The speed difference is less than or equal −3.5 r/min.

7 Conclusion

It is practicable for instantaneous speed method to diagnose single-cylinder misfire fault. The cylinder misfire fault will influence the engine's work ability, such as reducing the work peak of instantaneous speed. Speed difference can be used to describe and judge the instantaneous speed fluctuation.

References

1. Cao H (2005) Intelligent technology of fault diagnosis for diesel engine. National Defence Industry Press, Beijing
2. Sun Y (2004) Research on intelligent diagnosis technology based on instantaneous speed for diesel engine. In: Naval University of Engineering, Wuhan
3. Pu L (2000) Research on monitoring and fault diagnosis technology based on instantaneous speed for diesel engine. In: Wuhan University of Technology, Wuhan
4. Yu Y (2007) Research on monitoring and diagnosing for marine engine based on instantaneous angular speed and thermal parameters. In: Wuhan University of Technology, Wuhan
5. Cheng Y, Hu Y (1999) Estimation of combustion pressure distinction between cylinders Using Transient Speed of Flywheel. Transactions of CSICE 17(1):82–85
6. Wang J, Lu Q (2001) Research and improvement of transient-speed measuring methods. Veh Engine 132(2):31–34
7. Wu H, Jin D, Cai Y, Yang J (2006) Application research of fault diagnosis on stirling engine using instantaneous speed signals. Diesel Engine 28(6):38–40
8. Yang J, Pu L, Wang Z, Zhou Y (2001) Fault detection in a diesel engine by analysing the instantaneous angular speed. Mech Syst Signal Process 15(3):549–564
9. Wang Z (2004) Research on fault diagnosis of diesel engine using pattern recognition method. Wuhan University of Technology, Wuhan

Design and Implementation of Integrated Monitoring and Diagnosis System for Marine Diesel Engine

Nao Hu, Jianguo Yang and Yonghua Yu

Abstract Marine diesel engine is the main power source of ship. Its safety and reliability can be improved by monitoring and diagnosing engine's running status. A kind of Integrated multi-method multi-parameter Monitoring and Diagnosis (IMD) system for marine diesel engine is presented in the paper for the purpose of meeting the application in real ship environment. Test and information technology are fully reflected in the design of IMD system which can monitor the running status and diagnose a variety of common faults of marine diesel engine. IMD system contains six sub-monitoring systems which are Thermal Parameter Monitoring (TPM) system, Instantaneous Speed Monitoring (ISM) system, Cylinder Pressure Monitoring (CPM) system, Shaft Power Monitoring (SPM) system, Valve Leakage Monitoring (VLM) system and Piston Ring Monitoring (PRM) system.

Keywords Marine diesel engine · Integrated monitoring and diagnosis (IMD) system

1 Introduction

The marine diesel engine is the main power source of a ship. Once it break down, it would bring huge threaten to the ship and cause detrimental effect on the ship's operational efficiency since it has complex structure, interactional components and works in a severe circumstance. Therefore, it is imperative to research on the monitoring and diagnosis system which can master the engine's running states to

N. Hu (✉) · J. Yang · Y. Yu
School of Energy and Power Engineering, Key Laboratory of Marine Power Engineering and Technology Ministry of Communications, Wuhan University of Technology, Wuhan 430063, China
e-mail: hunaofly@163.com

J. Yang
e-mail: jgyang@whut.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_56

avoid the tremendous accident. There are some sorts of foreign monitoring and diagnosis systems such as MAN B&W Company CoCos-EDS system, EUB research Institute EUB-CDS system, Kyma Company Kyma Diesel Analyzer system and so on are now available in the market. These systems are mature in technology but have downsides of high prices. The domestic ones have defect of single-function, they can diagnose a certain fault without locating it precisely. Thus it takes a long period to troubleshooting. An integrated monitoring and diagnosis system for marine diesel engine is on researching to solve the problems above. Except the combination of software and hardware, the automatic test technology and computer information technology are fully used in the design, which allows it to fusion analysis and process a variety of faults together to realize the monitoring and diagnosis of marine diesel engine in all aspects.

2 Integrated Monitoring and Diagnosis Principle

2.1 Thermal Parameter Monitoring and Diagnosis Principle

Thermo parameter method uses engine working medium and engine operational parameters to monitor and diagnose engine components or the whole engine's working conditions. Engine working medium include air, combustion gas, lubricating oil and engine coolant, etc. the working conditions and status of lubricating oil, fuel oil, coolant, inlet air and exhaust, turbocharger system affect directly or indirectly the engine combustion temperature, pressure, power output and efficiency. Thus, the working condition and status of engine components and systems can be monitored and analysed through the thermo parameters of an engine. As the development of digitization and information technology, those engine thermo-parameters nowadays can be measured and monitored by sensors installed on the engine. Therefore, online monitoring and diagnosis of engine faults is practical now.

Figure 1 shows the logics of thermo-parameters method. A, B, C and F are defined as set functions, where $A = \{a_1, a_2, a_3, \ldots, a_n\}$ represents measurement parameters; $B = \{b_0 = 0,\ b_1 = 1\}$ is type of diagnosis. When $B = 0$, it indicates direct diagnosis, is able to reflect a fault directly. When $B = 1$, it represents indirect diagnosis, requiring an integration of several parameters to diagnose faults; $C = \{c_1, c_2, c_3, \ldots, c_m\}$ represents fault characteristics; $F = \{f_1, f_2, f_3, \ldots, f_k\}$ is identification of faults set which is corresponding to a single parameter or multiple parameters.

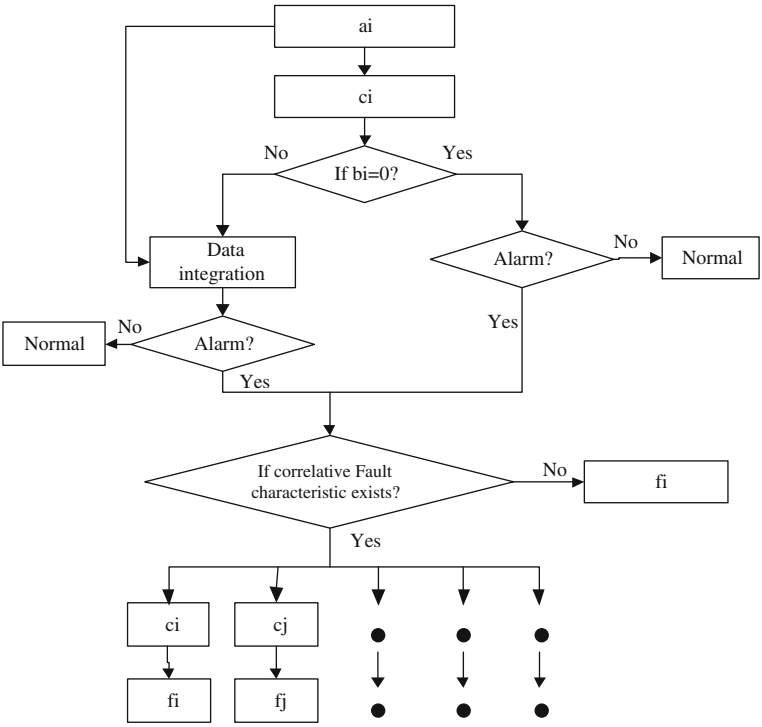


Fig. 1 Principle of thermo-parameter method

2.2 Instantaneous Speed Monitoring and Diagnosis Principle

From diesel engine working principles, the change of engine instantaneous speed is mainly due to the resultant force of cylinder gas force, engine reciprocating force and the total shaft torque, thus, for a multiple-cylinder engine [1], there is,

$$\left[I_0 + \sum_{i=1}^{n} mR^2 f_1(\theta_i)\right] \frac{d\dot{\theta}}{d\theta}\omega = \sum_{i=1}^{n} \left[p_i AR - m\omega^2 R^2 f_2(\theta_i)\right] f_1(\theta_i) \tag{1}$$

where, θ is engine crankshaft angle; pi represents cylinder pressure; A is piston cross-section area; R is radius of crankshaft; I_0 is the moment of inertia of shaft and fly wheel; $\dot{\theta} = \omega$ is shaft instantaneous angular speed (IAS); m represents the reciprocating mass; λ is ratio of crank radius to connecting rod length; n is number of cylinder.

$$f_1(\theta) = \sin\theta + \frac{\lambda \sin 2\theta}{2\sqrt{1-\lambda^2 \sin^2\theta}}; \tag{2}$$

$$f_2(\theta) = \cos\theta + \frac{\lambda \cos 2\theta}{\sqrt{1-\lambda^2 \sin^2\theta}} + \frac{\lambda^3 \sin^2 2\theta}{4\sqrt{\left(1-\lambda^2 \sin^2\theta\right)^3}}; \tag{3}$$

Instantaneous speed ω is mainly relative to cylinder pressure. Speed fluctuation reflects the working balance in an engine working cycle and is sensitive to the faults which result in working ability decline. Therefore, diesel engine running state can be monitored online by researching on instantaneous speed and its fluctuation rate.

2.3 Cylinder Pressure Monitoring and Diagnosis Principle

The cylinder pressure of diesel engine contains a wealth of information related to the engine performance, it can reflect the diesel engine cycle working conditions. The area of cylinder pressure stands for how much work it does in a work cycle. The inner cylinder combustion process, heat release rate, heat transfer between gas and cylinder wall, inhale and exhale process and fault information can be researched through cylinder pressure. For example, the peak cylinder pressure can reflect injection timing and the quality of air and fuel mixing; the reasons for a reduced compression pressure may be due to gas leaking, insufficient of inlet air or valve leaking; a fluctuation in the peak pressure on an indicating diagram may be caused by a blockage of injection nozzle or other problems associated with nozzle [2, 3].

2.4 Shaft Power Monitoring and Diagnosis Principle

Torsion would occur when the marine shafting system is running with load. Torque and power can be calculated through measuring the torsion extent of shafting system. The running state of marine shafting system contains much healthy information. The matching relationship between the ship, engine and propeller can be mastered via measuring the shaft power in a variety of working conditions. When it is bad, engine can't reach the rated power or run over load heavily. It will affect ship's general energy efficiency and power plant reliability greatly. The operating efficiency, condition and performance of ship can also be estimated by shaft power, which provides proofs for the fault diagnosis of diesel engine.

2.5 Air Valve Leakage Monitoring and Diagnosis Principle

When the diesel engine's cylinder head is knocked by air valve, it radiates acoustic emission signal, which reflects elastic wave feature of cylinder head material after impacted. The common faults such as air valve leakage and abnormal air valve interval would lead to the acoustic signal generated by cylinder head changes with cylinder head impact force. Its frequency varies from several Hz to hundreds Hz. It is proved that acoustic signal contains abundant information and can be easily used in the situation of monitoring air valve condition [4].

2.6 Piston Ring Wear Monitoring and Diagnosis Principle

Contact free magneto-resistive sensors are used for measuring the wear of piston ring. Wear occurs due to the friction between piston ring and cylinder liner when the engine is running. Characteristic parameter value output of magneto-resistive sensor changes with the magnetic field intensity which is decided by wear extent when piston ring goes through magneto-resistive sensors. Base on output characteristic parameter value, wear extent can be estimated, so as to monitor piston ring state and make maintenance advice.

3 Integrated Diagnosis Model and Strategy Design

There are many sorts of monitoring and diagnosis methods for marine diesel engine. Each method has its adaption and limitation. How to make full use of their merits and avoid their downsides is the key problem. TPM monitors a variety of parameters, which have features of good qualities and wide scope. But it has some defects such as thermal parameter can easily be affected by the interactions among cylinders and can't be used for recognize engine state when many different kinds of faults occur at the same time. IPM method is convenient to use. Its signal can be easily measured and fault criterion has Strong Commonality. However, its measure precision is influenced greatly by flywheel tooth indexing accuracy and sensor sampling frequency. Although it reflects the fault information, it can't diagnose out the reasons which cause the decline of cylinder power output. CPM plays an important role in researching the combustion process, combustion release rate and etc. However, because of the hostile working circumstance inside cylinder, CPM method measurement can't be continuous monitoring for a long time. SPM can monitor torque and power output of diesel engine on line in a long period, but it only can appraise the diesel's overall condition without figure out the reasons. VLM use acoustic emission signal to monitor the state of air valve. Acoustic emission signal can be recognized in an early stage of material damage. According to its

features and intensity, not only its source present state can be inferred, but also its history can be revealed and growing trend can be predicted. In an actual monitoring situation, there are many disturbs in acoustic emission signal due to the complex surroundings, which will leads to some misjudges. PRM use contact-free magneto-resistive sensors to measure the wear state of piston ring. This method has an inconvenience that it has to punching a hole on cylinder wall to install the sensors. The hole usually on the position where influence least to the tightness of combustion chamber.

From what have discussed above, select the proper monitoring and diagnosis methods to be adopted together according to the demand of real marine diesel engine can make a more precise judge of faults so as to improve ship economy and safety.

3.1 Integrated Monitoring and Diagnosis Strategy

Integrated monitoring and diagnosis system strategy is shown in Fig. 2.

IMD system adopts modular mounting and using style, integrates with off-line and on-line means to realize integrating diagnosis. The detail flow path is as follows: when engine works in a stable condition, in the first place, TPM,IPM and SPM methods are used for monitoring the whole engine's running state. To be more precise, TPM method monitors some components of engine and running condition of system; ISM method monitors and diagnoses diesel engine's dynamic balance while SPM method monitors the change of engine power output in real time. If faults were detected, CPM method is used for doing further diagnosis. It needs to measure a certain cylinder or all cylinders to locate the fault positions precisely. In order to reduce maintenance cost, VLM method and PRM method can be adopted. VLM method appraises each exhaust valve' state and provide diagnosis suggestion. PRM method estimates each piston ring in cylinder and thus used as basis for maintenance. All monitoring and diagnosis results are saved to database and managed efficiently.

4 Integrated Monitoring and Diagnosis System Structure Design

4.1 Hardware Design

Integrated monitoring and diagnosis system combines computer networks and database management technology. Its overall structure is shown in Fig. 3. It is mainly consist of engine room part and central control room part.

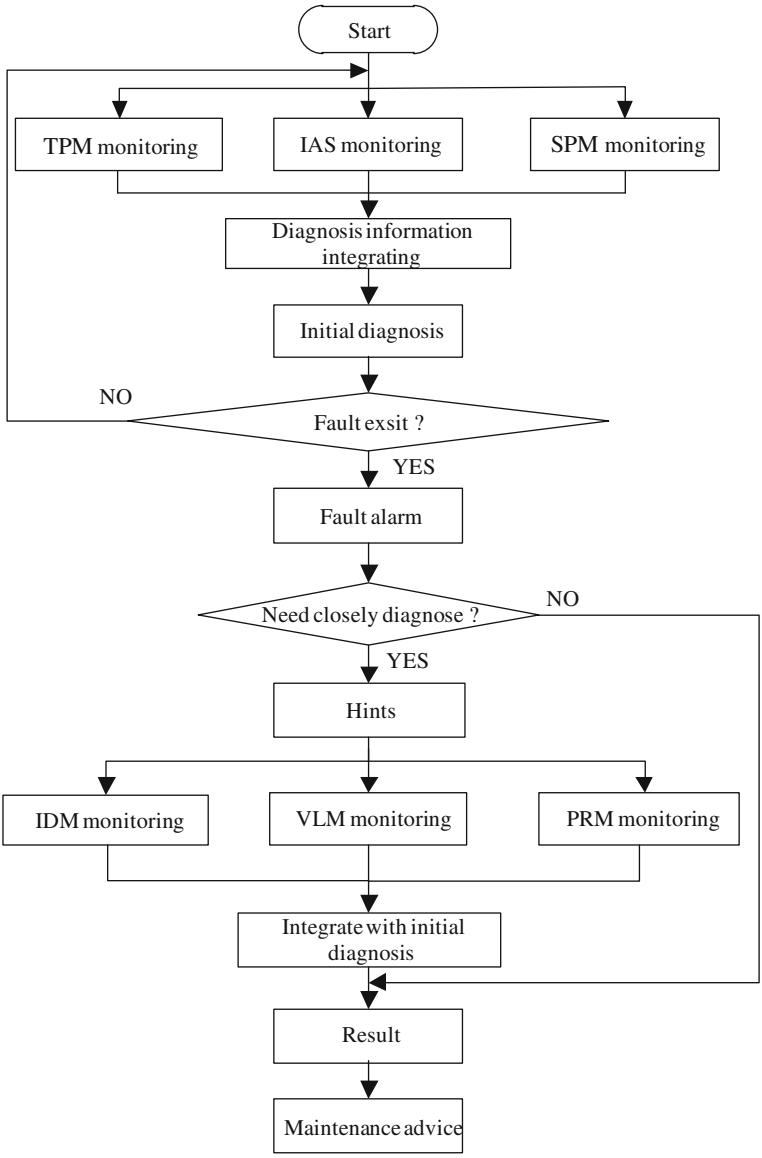


Fig. 2 System strategy

(1) Devices in engine room

A. Sensors: 2 magneto-electric speed sensors, which are used for measuring Top Dead Central (TDC) signal and Crankshaft Position (CKP) signal, shown as ① and ② in Fig. 2; Each cylinder has a magneto-resistive sensor used for piston ring monitoring, shown as ③. A cylinder pressure sensor is

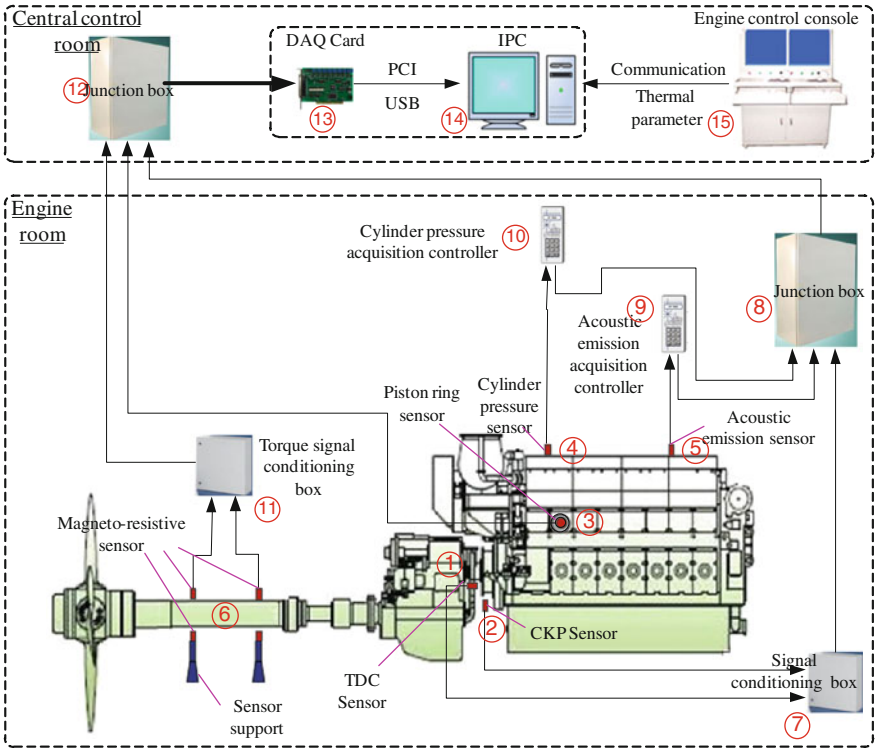


Fig. 3 System structure overall design

shown as ④. An acoustic emission sensor is shown as ⑤; Two couples of magneto-resistive sensors used for shaft power measuring are shown as ⑥.

B. Other devices: A TDC and CKP signal conditioning module ⑦ is mounted beside the diesel engine. It is used for changing sine wave signals outputted by two magneto-electric speed sensors into TTL pulse signals. The portable cylinder pressure measurement controller ⑩ is used for obtaining cylinder pressure signal in equal crank shaft interval and doing data process. Data is transmitted to industrial computer in central control room. The cylinder pressure acquisition system and associating sensor and cables are disconnected and stored after the measurement of cylinder pressure in connection box ⑧ nearby engine. The same rule is suitable to acoustic emission controller and its associating sensor and cables. Torque signal conditioning box ⑪ is used for conditioning the outputs of magneto-resistive sensors in order to meet the demand of AD card.

(2) Devices in central control room

Connection box ⑫ contains wire connector and DC module, which are used for integrating signals from engine room. Industrial computer ⑭ contains Ethernet

communication card used for obtaining thermal parameter and high performance data acquisition card used for measuring instantaneous speed signal, shaft power signal and acoustic emission signal. Besides that, a RS485 communication card is used for gathering cylinder pressure data.

4.2 Software Design

4.2.1 TPM Module

Thermal parameters are acquired via the engine room alarm systems normally transmitted with a series port connection or an Ethernet connection. Thermal parameter diagnosis is mainly according to diesel engine manual. Engine state can be judged and alarm can be generated via lube oil pressure and temperature, exhaust temperature, speed and etc.

4.2.2 ISM Module

Applying a PCI-7833R data collection card of a field-programmable gate array (FPGA), the IAS module measures the crankshaft angle pulse periodically to calculate the engine IAS according to Eq. 2. If the total number of tooth on fly wheel is Z, the engine IAS n_i can be expressed as,

$$n_i \approx \bar{n}_i = \frac{\Delta\varphi}{T_i} \mathrm{g} \frac{60}{360} = \frac{\Delta\varphi}{6\mathrm{T_i}} = \frac{360}{\mathrm{Z}} \mathrm{g} \frac{1}{6\mathrm{T_i}} = \frac{60}{\mathrm{ZT_i}} \tag{4}$$

where Ti is TTL pulse period (s), Z is the total number of teeth on fly wheel, Δφ is the crank angle (deg) corresponding to Ti. The measured data needs to be processed by digital filtering and cycle averaging before the calculation of the instantaneous angle speed is conducted.

Instantaneous speed fluctuation rate ε is adopted as criterion which is:

$$\varepsilon = \frac{v - \bar{v}}{\bar{v}} \times 100 \tag{5}$$

In formula: v stands for instantaneous speed, $\bar{v}$ is average speed.

Instantaneous speed fluctuation reflects on the deviation of instantaneous speed and average speed. Its absolute value is meaningless. When a certain cylinder has oil leak fault, instantaneous speed fluctuation rate ε declines distinctly. Based on that, instantaneous speed fluctuation rate ε is used as characteristic parameter to judge the fault of cylinder and find the fault cylinder number according to fire-order.

4.2.3 CPM Module

The module processes the measured TDC and cylinder pressure data by averaging the data of multiple cycles, data shifting to make the crankshaft signal with an equal-interval and data smoothing. Combined with engine structure data, the module evaluates engine performance by a synthetic analysis of the engine indicating pressure, compression pressure, expansion pressure, pressure increment, peak combustion pressure and indicating power calculated from the P − φ diagram [5].

Module of indicating diagram diagnosis: The measured indicating diagram is analyzed off-line by comparing with a troubleshooting list. At the same time, the analysis also incorporates with the measured thermal parameters. The following criteria have been used in identifying the faults: p_m—max. peak combustion pressure, it reflects the quality of air and fuel mixing and the correctness of injection timing; p_d—pressure at 5° before top dead center (BTDC), indicating the ignition timing; p_c—compression pressure, it can be used for identify problems associated with gas leaking; V-p_m—mean square deviation of pressure from 5° BTDC to 5° ATDC, it can show the pressure fluctuation at top of the indicating diagram; Tr—exhaust temperature, reflecting the slope of expansion line and post combustion phenomena. This module is able to identify some common faults by comparing the measured data with a model data base.

4.2.4 SPM Module

A couple of magnetic steel belts of same distribution are mounted at two sides of drive shaft. Magnetic steel belts rotate with drive shaft and the frequency of magneto-resistive sensor output changes with speed. Signal phase relationship is constant in two measure points and only affected by the relative installation position of magnets and magneto-resistive sensors. In another word, the initial θ_0 is a constant value. Drive shaft transforms due to torsion when engine is working with load. Measure the relative torsion angle θ between point A and point B. Its torsion angle is $\Delta\theta = \theta - \theta_0$. From formula (6) we can know that torque can be calculated via $\Delta\theta$. Combined with speed calculated through point A or point B, shaft power can be obtained.

$$T = \frac{G(D^4 - d^4)}{584L} \cdot (\theta - \theta_0) \tag{6}$$

According to the torque and speed, shaft power P can be calculated:

$$P = \frac{T \cdot n}{9550} \tag{7}$$

In formula, θ_0 is initial phase angle (°), θ_0 is initial phase angle (°), D is shaft external diameter (m), d is shaft inner diameter (m), L is the distance between two belts section(m), T is torque (N.m), P is shaft power(kW), n is speed (r/min).

FPGA board card PCI-7853R is used for measuring the signals of two couples of magneto-resistive sensors. Before collecting data, zero setting must be done. Phase difference method is adopted to extract time difference of two couples of sensors output. Base on principle above, the engine output power is monitored and recorded on line in real time. Combined with history data, it can analyze fault development trend.

4.2.5 VLM Module

FPGA board card PCI-7853R is used for measuring and analyzing the leakage of air valve. When acoustic emission signal are collected, time domain analysis and wavelet packet analysis methods are used for analyzing and processing it. Time domain analysis method obtains information of acoustic emission source through the parameters of acoustic emission sensor output. Its characteristic parameters include peak to peak value, RMS value, ring numbers, event count and energy. As a general rule, diesel engine fault can be diagnosed via comparison of acoustic emission signals in normal state and in fault state. But sometimes many kinds of fault signals superposition, which will impede the fault diagnosis via acoustic emission. Because acoustic emission signal has features of high time and space resolution and wide frequency band. Wavelet packet decomposition is adopted to do the multi-level classification of signal. High-frequency signal is decomposed to each frequency band so as to extract and analyze the signal including fault characters [6, 7].

4.2.6 PRM Module

FPGA board card PCI-7853R is used for monitoring piston ring signal of each cylinder. Time domain analysis method is adopted to analyze collected data. Sensor output signal amplitude has a corresponding relative with piston ring wear. Thereby, its amplitude can be used as characteristic parameter of piston ring wear. Data process is divided into two steps. First, signal population mean is removed by DC component de-mean; second, magnetic field shift is eliminated via filtering. In order to reduce random error, piston ring wear signal during 0°CA–180°CA in a work cycle is chose to do equal crank angle process. The data processed is stored into data base for the purpose of analyzing piston ring wear trend.

4.2.7 Display and Print Module

The main screen displays engine running state, diagnosis results and alarms. TPM module displays thermal parameter values and alarms. ISM module displays the speed fluctuation rate and maximum speed deviation. CPM module displays the main performance parameters of each cylinder in list and histogram comparison, and also P–V, P − φ diagrams. SPM module displays the output power of marine propelling engine and analyzes the trend development by comparing to history data. VLM module indicates air valves working conditions while PRM module displays piston ring signals and history data comparison. The results can be reported in a word file for crew to check and fill except TPM module. Software overall structure design is shown in Fig. 4.

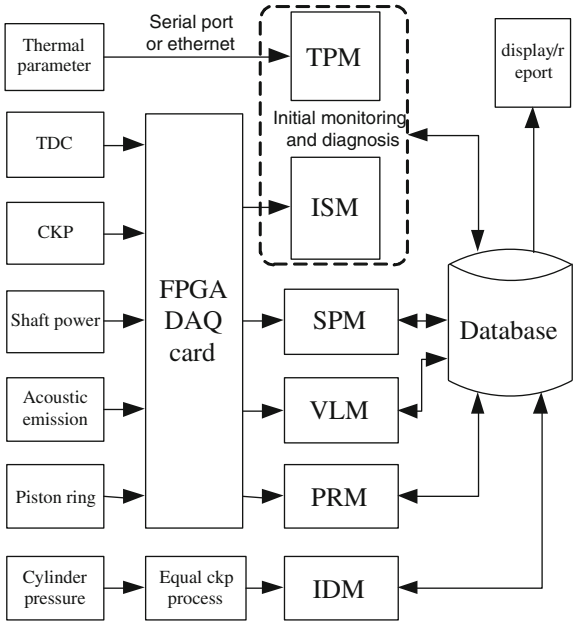


Fig. 4 Software overall structure design

4.2.8 Database Module

Based on the overall consideration of store and query efficiency, data size and operating stability, SQL2000 platform is adopted. It realizes data store and analysis via establishing basis information table, measurement data table, diagnosis state table, criterion table and etc. Data management method is divided into normal state

and fault state. Temporary data form and historical data form are used during data management. Temporary form only saves data in 2 h where data was written automatically every 15 min and managed in the way first in first out.

5 Conclusion

1. The diagnosis strategy is designed while marine diesel engine fault recognition rate and trouble-shooting accuracy are improved efficiently with multi-parameter multi-method integrating technology, virtual instruments technology and data base management technology.
2. Modular Design is used in every monitoring system, which has Strong adaptability and flexibility and can be optional selected according to monitoring and diagnosis demand of a variety of marine diesel engines.
3. The integrated monitoring and diagnosis structure is designed in detail. The feasible plan of applying to real ship is analyzed. It can monitor the fault of engine in time and improve ship operating economy.

References

1. Yonghua Y, Jianguo Y (2007) Design and implementation of the remote fault diagnosis system for 300 m^3 Dredger. J Wuhan Univ Technol 31(2):195–201 (Transportation Science and Engineering)
2. Lamaris VT, Hountalas DT (2010) A general purpose diagnostic technique for marine diesel engines—Application on the main propulsion and auxiliary diesel units of a marine vessel. In: Energy conversion and management, pp 740–753
3. Rongming D (1997) Marine diesel engine. Dalian Maritime University Press, Dalian
4. Douglas RM, Steel JA, Fog TL (2006) On-line power estimation of large diesel engines using acoustic emission and instantaneous crankshaft angular velocity. Int J Engine Res 7(5):399–410
5. Yonghua Y, Jianguo Y (2009) Development of performance monitoring analyzer for marine diesel engine. Ship Eng 32(1):8–11
6. Kurtis G, Ahsan K (1999) Application of wavelet transform in earthquake, wind and ocean. Eng Struct 21(3):149–167
7. Gary GY, Kuo CL (2000) Wavelet Packet Feature Extraction for Vibration Monitoring. IEEE Trans Ind Electron 47(3):650–667

Universal Wireless System for Bridge Health Monitoring

Mehdi Kalantari Khandani, Farshad Ahdi, Amirhossein Mirbagheri, Richard Connolly, Douglas Brown, Duane Darr, Jeffrey Morse and Bernard Laskowski

Abstract Since August 2010, Resensys wireless SenSpot tilt and strain sensors were deployed to monitor a highway bridge in Maryland. Similar installations were performed in bridges in the US, Canada, and Indonesia. Signal analyses concluded important observations about the response of the bridge bearings to change of temperature. In some instances, the change in the strain exceeded 30 microstrains, e.g., due to the bridge rehabilitation work. The decision parameters about the state of a structure were fused to produce the structural integrity knowledge to be used for predictive diagnostics. Using this method, the monitoring system can predict rupture, crack, yielding or generally any signals before the collapse of a structure or any member damage before it happens. Finally, Resensys is working with Analatom, Inc, and other third-party OEMs in the development of an Onboard SHM Data Aggregator Module (OSDAM) platform to manage and control access to multiple SHM Sensor Systems.

1 Introduction

Deterioration of critical infrastructure—such as bridges, pipelines, buildings, and railways—is a common, yet complex problem. Currently, manual expert inspection is the most common practice used to monitor the structural integrity of bridges and other civil infrastructures. However, manual inspections have proven to be insufficient to ensuring safety. Such inspections do not provide enough information to prevent catastrophic failures. In the US only, the magnitude of the bridge safety problem has been highlighted by the Federal Highway Administration, which determined that 71,429 bridges in the United States are rated as structurally defi-

M.K. Khandani (✉) · F. Ahdi · A. Mirbagheri · R. Connolly · D. Brown · D. Darr · J. Morse · B. Laskowski
Resensys LLC, 387 Technology Dr, College Park, MD 20742, USA
e-mail: mehdi@resensys.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_57

cient [1, 2]—the same rating assigned to the Minneapolis I-35 W Bridge before its collapse [3]. The global scale of this problem is significantly larger.

Beyond manual inspections, other existing techniques for structural health monitoring suffer from non-scalability due to the high cost of instrumentation devices, large installation costs (e.g., due to wiring needs), or high maintenance costs. To ensure public safety and the continuous serviceability of bridges and other infrastructure systems, it is imperative to develop cost effective, easy to use, and reliable technologies that regularly assess their structural health and integrity.

2 Structural Health Monitoring Using SenSpot Sensors

To protect highway bridges and other high values structures, Resensys offers a solution that combines a number of recent and emerging technologies—micro-structured sensing, ultra-low-power wireless communication, and advanced microelectronics—into a novel, small, and light-weight wireless device known as SenSpot [4]. The SenSpot sensors offer high performance for large-scale sensing, wireless synchronization, and ultra-low-power wireless communication. Due to their small size and light weight, Resensys's SenSpot sensors can be applied easily to as many points on a structure as needed, with minimal installation effort. Additionally, the sensors are self-calibrated and execute self-diagnostics if necessary within a few seconds.

The SenSpot sensor is powered using a prime lithium-ion battery that is designed to supply the required energy. In most applications, such as the long-term monitoring of structures that do not need frequent sampling, the sensor is designed to work for several years (typically up to 20 years), due to ultra-low-power energy consumption. The device uses Resensys's proprietary sensing, synchronization, and ultra-low-power wireless communication technology. A picture of a conventional SenSpot installed on a bridge bearing is shown on the right of Fig. 1. In this installation, the sensor reports the vibration, inclination, and strain of a bridge bearing, and monitors changes in these parameters as the bridge expands or contracts in response to temperature variations.

Resensys SenSpot sensors have a number of unique features that significantly distinguish them from existing solutions for structural collapse prediction.

These features include:

Fast and easy installation: SenSpot sensors are very lightweight (40–50 g) and small size.

Low cost: The sensors are low cost (in volume, $40–50 per device); the low cost, combined with the ease of installation, enables their large scale use.

Long lifetime: Resensys's proprietary ultra-low-power sensing and wireless communication technology enables SenSpots to operate for relatively long time (up to 20 years).

Fig. 1 Resensys SenSpot (*left*) and installed sensor on a bridge bearing (*right*) for tilt and loading monitoring

Strong software tools: Resensys offers a complete solution: not only the SenSpot sensors but also the base display unit which can be attached to hand-held devices for interoperability. Also, a powerful software package analyzes the data to predict structure collapse time and display the results to incident commanders in real-time.

Currently, SenSpot sensors offer a variety of features that meet the needs for monitoring bridges: strain, tilt, inclination, moisture and humidity, vibration, temperature, pressure, instantaneous displacement, maximum/minimum displacement, crack activity, deformation, etc. A number of these SenSpot sensors are shown in Fig. 2.

In addition to the SenSpot sensors, a complete Resensys structural health monitoring system includes software and hardware components for (1) reliable collection of SenSpot data, (2) aggregation of data, (3) addition of timestamps, (4) communication of data to a remote server, and, finally, (5) data visualization and the identification of structural issues. (More details about the Resensys solution is provided in Sect. 4, Related Research/Research and Development).

In addition to the SenSpot sensors, a complete bridge health monitoring system based in the proposed approach includes software and hardware components for (1) reliable collection of SenSpot data, (2) aggregation of data, (3) adding timestamps, (4) communicating the data to a remote server, and, finally, (5) visualizing the data and detecting structural issues. Figure 3 shows a picture of a practical structural health monitoring system, which includes the following components:

Fig. 2 *Left* to *right*: Samples Resensys SenSpots sensors

SenSpot sensors attached to structure (average 10–100 per bridge, depending on design and monitoring needs).
A data collector (known as SeniMax), which collects data on site of SenSpot and sends it to a remote server (1 per bridge).
Software (known as SenScope) that analyzes data and generates alerts.

3 Bridge Monitoring Case Studies

SenSpot tilt sensors can gather changes in the tilt of the bearing by angles as small as 12 arc s (or approximately 0.003°). Tilt changes happen as a result of daily temperature variations. Such changes can be monitored by the installed tilt sensors. The installed tilt sensor on a rocker bearing is shown in Fig. 4.

The tilt data gathered by the sensor is wirelessly transmitted to a wireless data collector (SeniMax) at the bridge, which in turn transmits data wirelessly to a remote server for processing, visualization, and archiving. Installation is straightforward, and for best performance, it is recommended to install the sensors on the side of bearing as shown in Fig. 4. Figure 5 shows a bearing's tilt readout provided by the installed SenSpot during 24-h period from July 28th 2011 until July 29th 2011. As shown in the figure, the SenSpot reported tilt measurements ranging from 1.70 to 2.45°. In other words, the change in tilt of the bearing was 2.45−1.70 = 0.75°.

In this case, the theoretical tilt change can be calculated by considering the temperature variation of the mentioned day. During this 24-h period, the minimum temperature was 69 °F and the maximum was 89 °F. Therefore, the variation of the temperature that day was 20 °F.

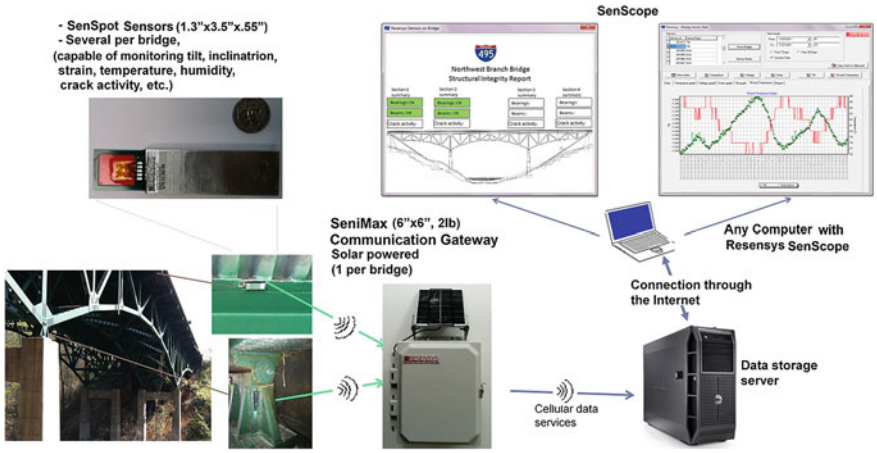


Fig. 3 Illustration of a complete bridge health monitoring system based on SenSpot sensors

Fig. 4 Wireless tilt sensor on a rocker bearing of a highway bridge

The main words in all headings (even run-in headings) begin with a capital letter. Articles, conjunctions and prepositions are the only words which should begin with a lower case letter.

Figure 6 shows the drawing of the rocker bearings on the bridge (from bridge design sheets). To calculate the dependence of the tilt of the bearing on the temperature, we use the following data:

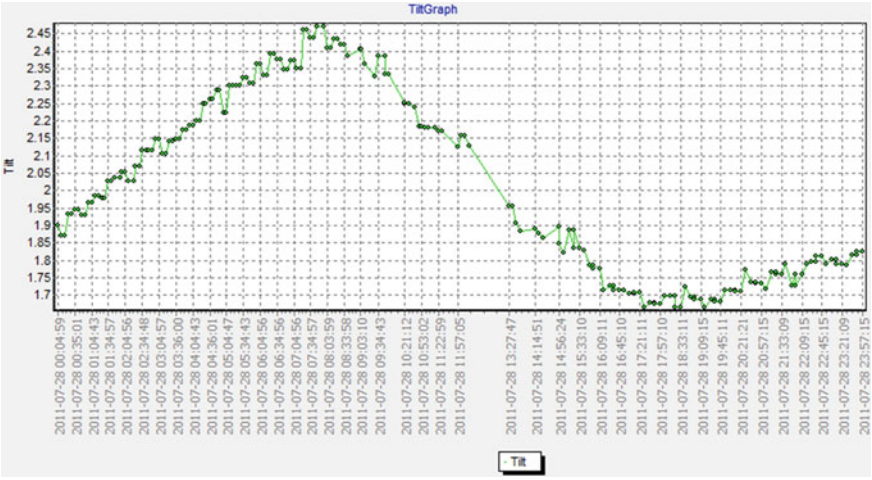


Fig. 5 Bearings tilt variation during 24-h period

R The radius of the bearing = 1′ 6″ = 18″ = 0.4572 m (shown in Fig. 6)
D the expandable portion of bridge deck = 150′ = 45.75 m
λ Thermal expansion coefficient of the bridge: 12×10^{-6}

Therefore, per degree Celsius change of temperature, the change in the expandable portion of the bridge is:

$$\mathbf{\Delta L = \lambda \times D = 12 \times 10 - 6 \times 45.75 = 0.00055\ m = 0.55\ mm}.$$

As a result, the change in the tilt per degree Celsius is:

$$\mathbf{\Delta\theta = tg - 1\,(\Delta L/R) = tg - 1(0.55\ mm/0.4572\ m) = 0.0012\ radian = 0.068^{\circ} = 248\ arc\ s}.$$

In other words, the theoretical change in tilt of the rocker bearing per degree Celsius is 0.068°, which is equivalent to 248 arc s. Equivalently, the amount of change in the tilt of the bearing per degree Fahrenheit can be calculated by dividing the above numbers by 8. Therefore, the theoretical change is 0.038° or 139 s per degree Fahrenheit. As a result, the theoretical tilt change is calculated to be 139 s change per degree Fahrenheit. Therefore, the theoretical change in tilt is calculated as: 139 × 20 = 2780 arc s.

To summarize, the theoretical and measured tilt can be compared as follows:

The measured tilt variation using SenSpot: **2734 s = 0.7594°.**
The theoretical tilt variation: **2780 s = 0.7722°.**
Error: **46/2780 = 1.6 %**.

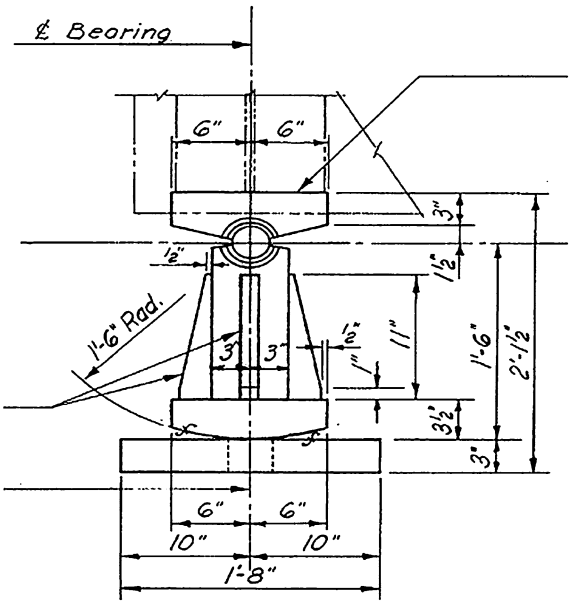


Fig. 6 The schematic of rocker bearings, per design sheets of the bridge

3.1 Bearing Change Versus Temperature

As an interesting observation, the graph in Fig. 7 shows the tilt of bearing versus temperature reported by the SenSpot sensor on the bearing during a 3-week period (from March 20, 2012 until April 9, 2012). The green graph in the figure shows tilt while the purple graph shows the deck's temperature measured by a different SenSpot. As shown in the picture, the change in temperature of deck affects the tilt of the bearing. An increase in the deck's temperature results in a decrease in tilt, and conversely decreasing temperature of the deck increases the tilt. The shown behavior indicates a healthy behavior of the bearing. As an interesting observation, the temperature of the deck during this time period was reported to be around 78–25 = 53 °F; therefore, the expected change in the bearings tilt should be 53 °F × 139 s/ °F = 7367 s = 2.04°. This expected change is consistent with the readout of the SenSpot, which reported approximately 2° of change in the bearing tilt (from −1.0° to +1.0°) as can be seen in Fig. 7.

3.2 SenSpot for Strain Measurement on Bridge Deck

In this case, we study measurement of strain using the Resensys SenSpot on a bridge deck. The installed SenSpot is shown in Fig. 8. Due to rehabilitation work, there were numerous instances where the sensor detected sudden shifts or increases

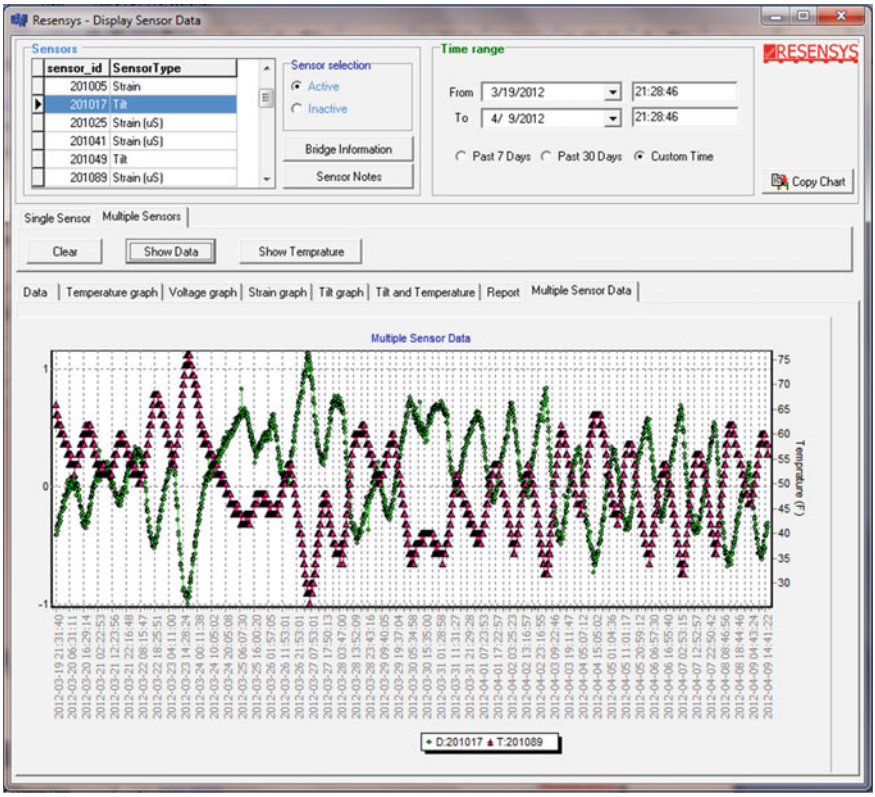


Fig. 7 A 3-week trace of bearing tilt measured by SenSpot versus temperature

Fig. 8 SenSpot under the deck of the Bridge

in strain readouts. In some instances, the change in the strain some exceeded 30 microstrains, which, under general conditions, could be a sign of significant structural change. However, all of the changes were revealed to have been caused by the rehabilitation work, such as the presence of heavy machinery on the bridge the placing of a large portion of concrete deck in median area (which was detected as a significant strain shift on October 31st 2011).

As one particular example of detection by the system, Fig. 9 shows an instance in which a steady increase in strain readout (by approximately 30 microstrains) was detected between Tuesday January 24th 2012 at 14:06 and Wednesday January 25th 2012 at 07:02. Generally, such an increase is an indication of severe change in loading pattern. In this case, it was confirmed that the increase in strain was due to the presence of heavy machinery on the bridge deck (as shown in the photo) during the mentioned hours. By using SenSpot data, the system detected many other events similar to this one, and all were verified with ground truth data.

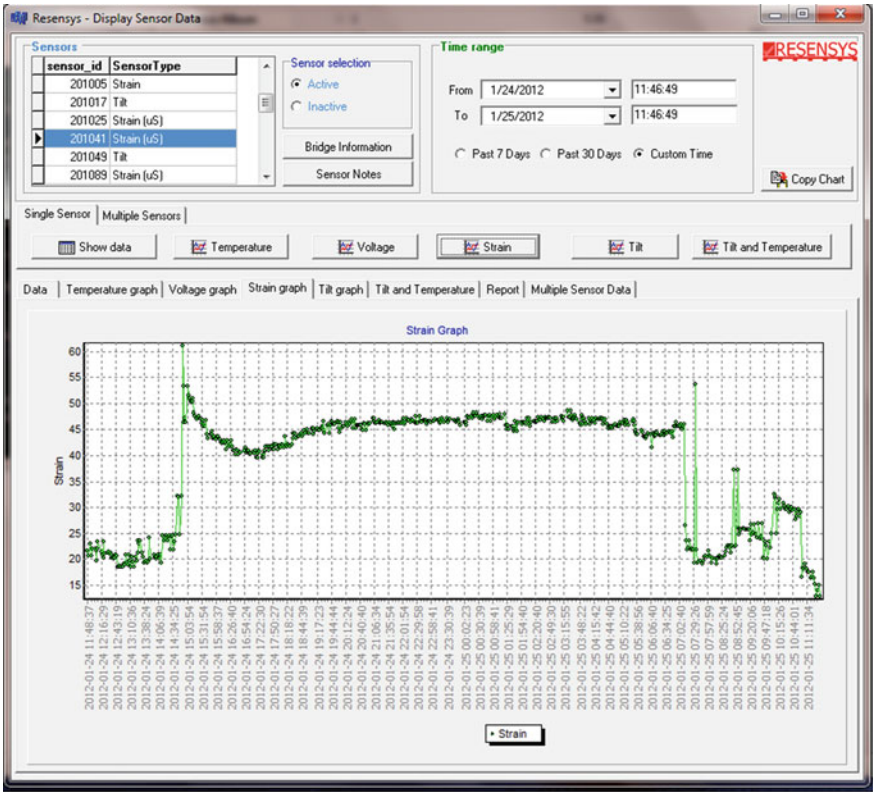


Fig. 9 SenSpot under bridge detects a sudden increase on 1/24/2012. The confirmed cause of increase was operation of heavy bridge rehabilitation machinery on the bridge deck

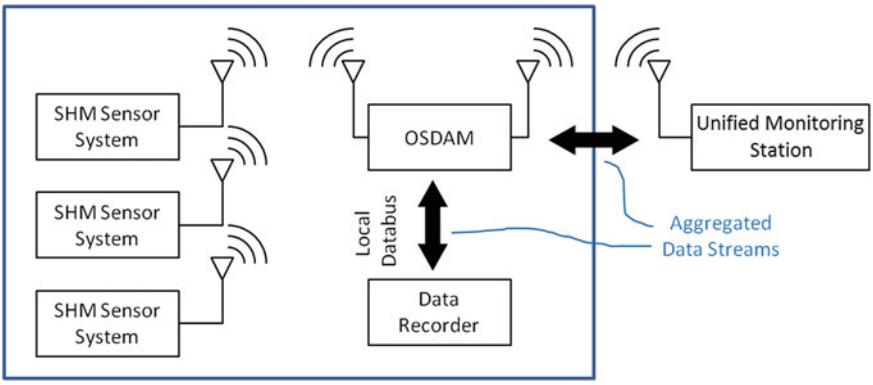


Fig. 10 Block diagram of OSDAM in proximity to the SHM sensor systems

4 Interoperability and Open Interface with Other SHM Systems

Flexible and universal access to structural health monitoring (SHM) data is important requirement of remote monitoring system. To achieve this goal, Resensys has been working with Analatom, Inc. and other third-party OEMs in the development of an Onboard SHM Data Aggregator Module (OSDAM) platform to manage and control access to multiple SHM Sensor Systems. OSDAM contains RS-485 and RS-232 wired serial interfaces combined with a dynamic, adaptive low-power ZigBee compatible wireless interface allowing its intelligent controller to coordinate, collect, and integrate multiple independent SHM sensor data streams [5]. Dynamic adaptive ZigBee interfacing provides OSDAM the unique capability to communicate with any ZigBee wireless sensor Personal Area Network (PAN) supporting the ZigBee PRO standard; thereby, enabling inter-PAN data transfer between independent ZigBee networks—a feature currently not supported by ZigBee standards. An example of an OSDAM in proximity to the SHM sensor systems is provided in Fig. 10. In this configuration, the OSDAM unifies the data streams for each SHM sensor system and distributes it to a local data recorder or ground station using a ZigBee wireless interface. The OSDAM connects the host to multiple SHM systems (up to 16 in total). To achieve this, a set of commands and responses are sent between the host and OSDAM and relayed through a set of commands sent between the OSDAM and SHM system. A common API was developed to communicate between the OSDAM and each SHM system to initiate remote data logging, real-time data logging, data download, and retrieve the system status. An example showing an Analatom AN110 SHM system connected wirelessly to a base station via an OSDAM is shown in Fig. 11.

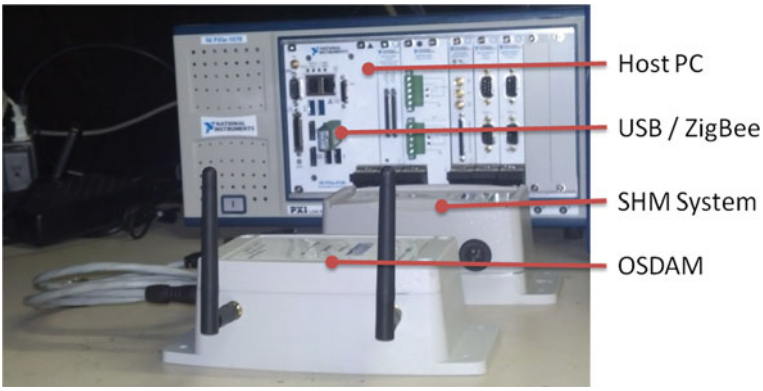


Fig. 11 Testing an Analatom AN110 with an OSDAM and a host PC

5 Conclusion

SenSpot sensors provide a wireless, easy to use, cost effective, and reliable solution for health monitoring of bridges and similar structures. The evaluation of SenSpot sensors on highway bridges showed that the devices provide a very good accuracy, a high reliability, and a good consistency in monitoring correct operation of bridge rocker bearings and loading on the bridges structural components. In addition to tilt and strain monitoring, other variations of SenSpot are available to monitor quantities such as moisture and humidity, vibration, temperature, pressure, instantaneous displacement, maximum/minimum displacement, crack activity, deformation, etc. SenSpot sensors and complete bridge health monitoring based on them are commercially available through Resensys LLC (www.resensys.com).

References

1. National Bridge Inventory (2010) US Department of transportation federal highway administration. http://www.fhaw.dot.gov/bridge/nbi.htm
2. Anderson GR Jr, Paul D, Kevin H, Jonathan K, Matt S, Matt S (2007) City Pages, Falling down. Village Voice Media, New York vol. 28 p 1392
3. Cohen S, Bakst B. (2007) Minn. Bridge problems uncovered in 1990. ABC News (abc.go.com). The Walt Disney Company, USA
4. Kalantari M (2011) New sensor technology enables self-powered, wireless structural monitoring. NACE International
5. ZigBee (2012) Wireless technology for low-power sensor networks: NACE International. www.commsdesign.com. Accessed 18 Oct 2012

Corrosion Detection on Buried Transmission Pipelines with Micro-Linear Polarization Resistance Sensors

Richard J. Connolly, Douglas Brown, Duane Darr, Jeffrey Morse and Bernard Laskowski

Abstract This paper presents an experiment adapting linear polarization resistance-based corrosion sensors, originally developed for aerospace applications, to measure the corrosion rate of API 5L ERW grade-B steel natural gas line pipe using micro-sized linear polarization resistance (μLPR) sensors made from the same alloy and grade steel. Sensors were installed under a 15 mil coating of fusion-bonded epoxy, at various proximities to a 1/8 inch defect introduced at a weld joint and along the pipe seam. After sensor installation the pipe was buried in an controlled environment with soil amended to a pH of five. This environment was held at a temperature above 35 °C while soil moisture content was modulated between wet and dry cycles, each lasting 7 days. LPR and environmental measurements were sampled at 5 min intervals. Post processing was performed to convert the LPR measurements to a surface-loss. Comparisons made in the data showed API 5L ERW grade-B steel natural gas pipelines were highly susceptible to corrosion along the seam, with all sensors showing activity in this region early in the experiment. Sensors adjacent to a weld joint began to display evidence of corrosion more slowly. These results verify the ability of μLPR sensors to measure corrosion activity under protective coatings in underground environments.

Keywords Corrosion · Diagnostics · Health management · Linear polarization resistance

R.J. Connolly (✉) · D. Brown · D. Darr · J. Morse · B. Laskowski
Analatom, Inc, Santa Clara, CA, USA
e-mail: richard.connolly@analatom.com

D. Brown
e-mail: doug.brown@analatom.com

D. Darr
e-mail: duane.darr@analatom.com

J. Morse
e-mail: jeff.morse@analatom.com

B. Laskowski
e-mail: bernard.laskowski@analatom.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_58

1 Introduction

Degradation of natural gas pipelines can result in costly maintenance and repairs, in addition to the presence of potential safety hazards. As such, there is a need to develop technologies in areas such as damage prevention, pipeline integrity, leak detection, and plastic pipe innovations. An efficient way to extend the life and lower maintenance costs associated with natural gas pipelines is to detect surface coating defects by measuring surface loss rates on corrosion sensors placed under coatings. Analatom, a structural health monitoring (SHM) company, has developed an innovative corrosion sensor system providing a holistic solution for monitoring high value structures. This system is low-cost, simple to install, and uses an easy to operate interface. It is built around a low-powered generic interface node capable of wirelessly communicating with a myriad of common SHM sensors. In this manner, structures can be fitted with sensors in any desired or designated location and format without the need for expensive communications and complex to route power lines.

The presence of oxidation reactions on natural gas pipelines requires corrosive species gain access to metal surfaces through protective organic coatings. Access to metal surfaces occurs through coating defects, such a mechanical damage or through erosion of protective coatings. This corrosion spreads to consume adjacent metals, where coatings remain intact. The mechanism responsible for this corrosive activity is diffusion of electrolytes, oxygen, and water molecules along the metal-coating boundary [1, 2]. This concentration driven process transports corrosive species along the interface causing surface metal loss and further separation of the organic coating. This process represents a significant cost for corrosion control and integrity management programs in the natural gas pipeline industry. Presently there are over 528,000 km of natural gas pipelines, 119,000 km of crude oil transmission pipelines, and 132,000 km of hazardous liquid transmission pipelines. This represents a consolidated infrastructure investment of $63.1-billion for all natural gas companies [3]. Replacement cost of these transmission and gathering pipelines is estimated to be approximately $643,800 per km with corrosion contributing as the primary factor in controlling asset life. Annual corrosion-related costs are estimated at $7.0-billion. Typically, corrosion operation and maintenance cost natural gas companies between $3,100 to $6,200 per km, which costs the industry between $2.4-billion to $4.8-billion [4]. Major pipeline companies have indicated these corrosion related cost are commonly due to cathodic protection failure resulting from coating deterioration or inadequate cathodic protection current.

SHM systems aim at reducing the cost of maintaining high value structures through the application of condition based maintenance (CBM) schemes [5]. These systems must be reliable, low-cost, and simple to install with a user interface designed to be easy to operate. To reduce the cost and complexity of such a system a generic interface node that uses low-powered wireless communications has been developed. In this manner a structure such as a bridge, aircraft, or ship can be fitted with sensors in any desired or designated location and format without the need for

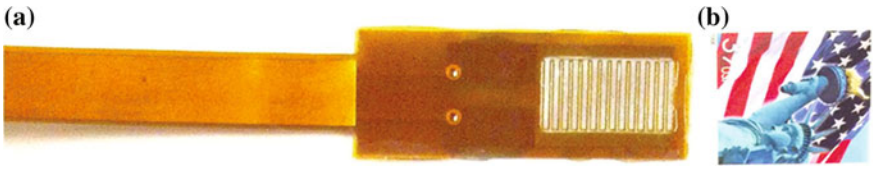


Fig. 1 Comparison of (**a**) the μLPR sensor with a flex-cable and (**b**) a postage stamp

communications and power lines that are inherently expensive and complex to route. Data from these nodes is transmitted to a central communications personal computer (PC) for data analysis.

Corrosion monitoring under coatings has been accomplished with Analatom's micro-sized linear polarization resistance (μLPR) sensor. This direct measurement technique utilizes a small form-factor sensor, shown in Fig. 1, to provide real-time measurements of metal loss and corrosion rate in remote and hard-to-access areas. A variety of methods are capable of experimentally determining instantaneous LPR such as potential step or sweep, current step or sweep, impedance spectroscopy, as well as statistical and spectral noise methods [6]. Analatom's SHM system uses the potential step (or sweep) approach to measure LPR. Direct measurements of LPR using the potential step method can be used to indirectly measure the corrosion current occurring between the two electrodes. Relating this to Faradays' Law with knowledge of the oxidation-reduction reactions and material properties, the amount of accumulated mass loss, and thereby surface loss, can be inferred.

2 Methods and Procedures

Analatom μLPR sensors were used to monitor environmental corrosion occurring on a buried section of line pipe. To perform this feasibility experiment a section of pipe was obtained, coated with fusion-bonded epoxy, sensors were installed at predetermined locations, defects were applied to the protective coating, and the pipe was placed in an environmental chamber. Accelerated corrosion was achieved by modulating temperature, moisture content, and pH of the soil. LPR data were collected and analyzed to determine the corrosion rate and surface loss that had occurred.

2.1 Pipeline Materials and Coatings

To perform this experiment a 3-ft section of API 5L ERW Grade-B D.R.L. schedule 40 steel line pipe with a 6-in diameter was utilized. The inside of the pipe was sealed from the external environment with welded end caps. After obtaining this pipe, it was coated with 15-mils of fusion-bonded epoxy.

2.2 Sensor Fabrication and Attachment

Analatom's μLPR sensors were fabricated from a section of API 5L Grade-B steel line pipe, which was cut down to make shim stock. Sensors were chemically etched to produce 300 μm wide interdigitated electrodes with a 300 μm spacing between electrode pairs and thickness of 10-mils. After etching, sensors were placed on a sheet of 2-mil Kapton film with 1-mil of adhesive.

After fabrication, sensors were mounted onto 18-in long flex cable assemblies and installed on the coated test pipe. Sensors were installed at eight predetermined locations: three sensors were installed adjacent to a weld joint; three were installed on the pipe seam; and two were installed on the midsection of the pipe. To attach each sensor, fusion-bonded epoxy was first buffed off of 2-in2 rectangular areas to expose the bare pipe. Industrial strength epoxy was applied to uncoated areas and μLPR corrosion rate sensor assemblies were attached to the pipe at room temperature. After curing the areas were sealed using 3 M Scotchkote Liquid Epoxy 323 at room temperature. This process is illustrated in Fig. 2.

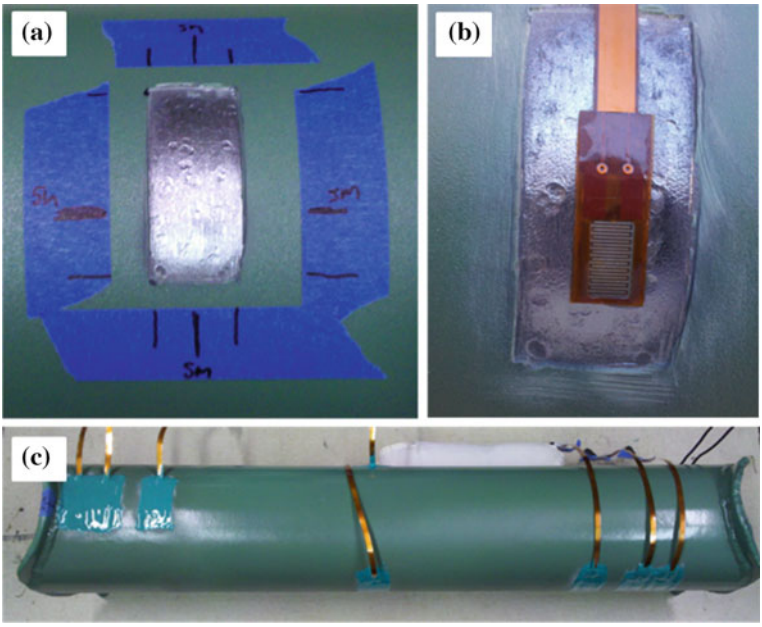


Fig. 2 **a** Area where FBE was buffed off to attach corrosion sensors, **b** μLPR sensor attached to the pipe surface with industrial strength epoxy, and **c** sensors coated with field patch epoxy material

2.3 Defect Installation

Blemishes were placed in the field patch material bordering two sensors to initiate corrosion. These defects were installed with a 1/8-in drill bit and completely removed the field patch material to expose the underlying bare pipe. Defects sites were chosen along a weld joint and on the pipe seam, which are both areas highly susceptible to oxidation. In both cases these discontinuities were located 3/8-in away from the midline of the nearest sensor. The next sensor in the immediate location of the defect was 1.25-in away. The final sensor in the vicinity of the defect was 3.75-in inches away from the sensor adjacent to the 1/8-in hole. Figure 3 illustrates the location of the defect relative to the surrounding sensors.

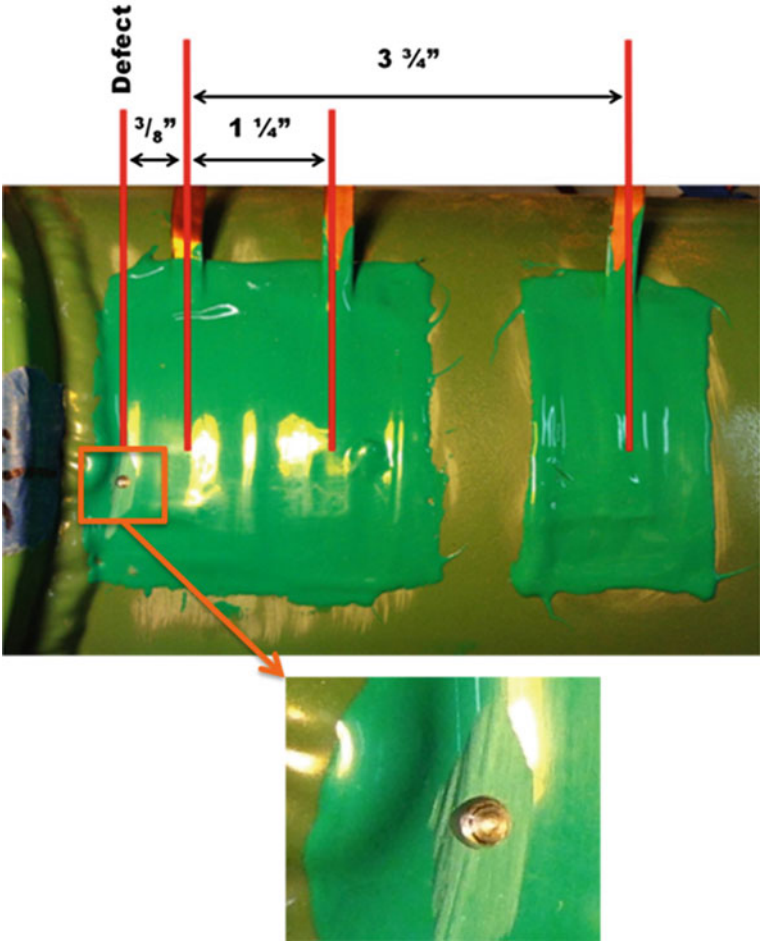


Fig. 3 Defect location relative to nearest sensors

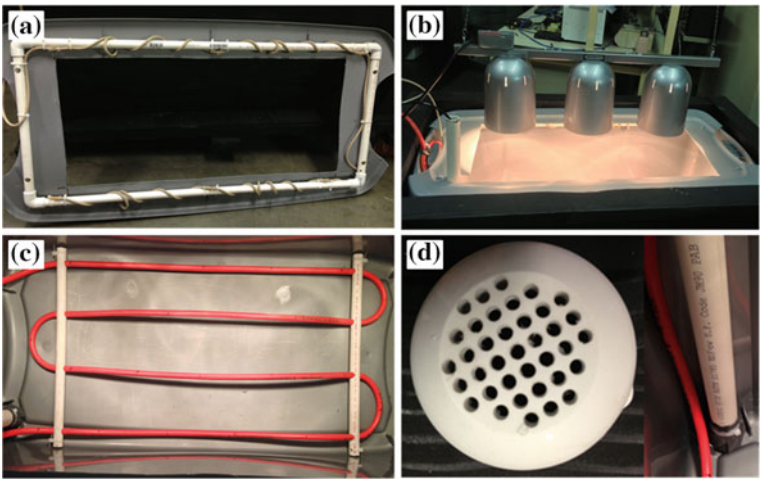


Fig. 4 The environmental chamber contained (**a**) wetting, (**b**) heating, (**c**) air sparging, and (**d**) chemical recycling subsystems

2.4 Environmental Chamber

Accelerated corrosion testing required fabricating an environmental chamber capable of efficiently changing the conditions experienced by the buried pipeline. A 50-gal chemically resistant plastic container was utilized to contain the soil and section of pipe. After having a suitable container for the system, a wooden exoskeleton was fabricated to provide structural rigidity and support for the heating system, air sparging system, wetting components, and chemical recycling system. These subsystems are shown below in Fig. 4.

2.4.1 Wetting Subsystem

Increasing the soil moisture content was achieved with a misting nozzle assembly attached to the bottom of a modified tub lid. This system was plumbed to a 1/4-in water supply spigot through a normally closed water supply solenoid. A programmable multi-event digital timer controlled the solenoid through a power interface adapter.

2.4.2 Heating Subsystem

A 750 W infrared heat lamp system was used to increase the temperature of the environmental chamber. The infrared lamp fixture was mounted above the exoskeleton and contained three 250 W infrared lamps. These lamps were connected to

a programmable multi-event digital timer through a 5 V power controller signal interface adapter. In addition to providing heat for the soil this system was an integral component of the drying cycle, along with the air sparging system.

2.4.3 Air Sparging Subsystem

To accelerate soil drying a system was installed to flow dry air through the soil to stripping away moisture. This sparging system was constructed from a 1/4-in diameter air hose mounted to a 1/2-in PVC pipe frame at the bottom of the environmental chamber. Perforations measuring 1/16-in were placed at 6-in intervals along the length. Compressed air was provided at 10 l/min. Airflow was actuated by a programmable multi-event digital timer connected to a 120 V normally closed 2-way solenoid valve.

2.4.4 Chemical Recycling Subsystem

A sump system to facilitate chemical recycling was fabricated and installed in the plastic container. This sump system was constructed from a PVC pipe with a 1-in diameter. A standard 1-in PVC cap was perforated with 37 holes with a 1/8-in diameter and affixed to the end of the pipe. To prevent these holes from clogging two layers of a mesh material were wrapped around the bottom of the cap to act as filters. The outermost layer was wrapped with polyester screen and the inner layer with spunbound polypropylene. Water containing leached chemicals was extracted from this pipe by aspirating with a vacuum pump.

2.5 Environmental Chamber Operation

Once the subsystems were installed, the pipe was placed inside the container and buried with amended soil. To increase the corrosion potential 1 g of aluminum sulfate was added per liter of soil, which lowered the pH to 5. Prior to the experiment, and weekly after starting, the soil pH was verified with a soil acidity tester. Figure 5 shows the orientation of the sensors and defect locations when the pipe was placed in the environmental chamber.

Operational conditions for accelerated corrosion testing during this experiment were selected based on modifying soil moisture content and temperature. At the beginning of the experiment soil moisture content was held between 40–50 % for a period of 7 days and then dried down to 25 % over a 7 day period. As the experiment progressed it was discovered these conditions were inadequate to initiate corrosion activity on the sensors under the fusion-bonded epoxy coating. After this time the temperature of the soil was raised with the infrared heat lamp system and the moisture content remained elevated.

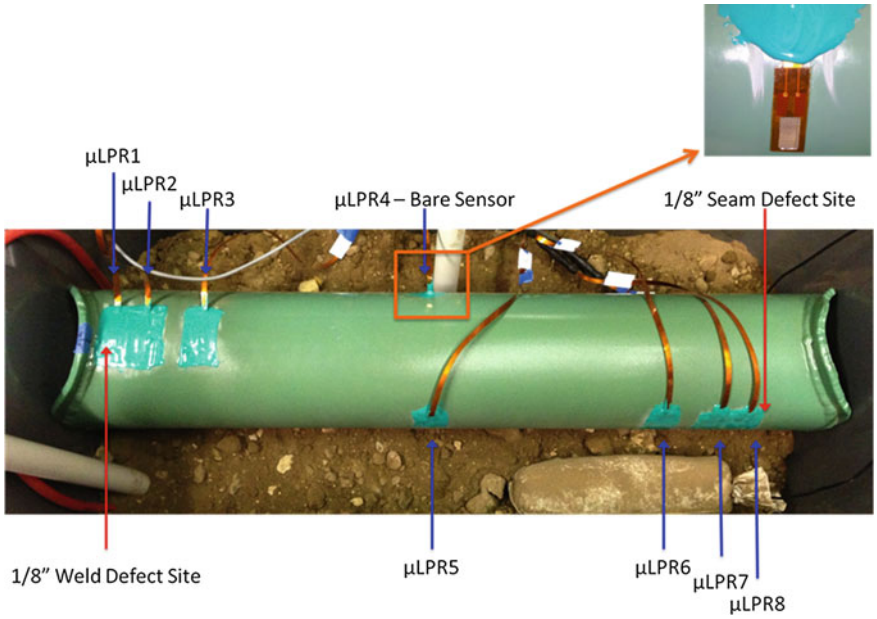


Fig. 5 Orientation of pipe section in the environmental chamber

2.6 Data Collection and Analysis

The Analatom AN101 corrosion monitor and data logger shown in Fig. 6 was used to monitor and record corrosion rates of the eight μLPR sensors on the test pipe. In addition to LPR data, the unit recorded relative humidity and temperature directly above the soil. Each AN101 node consists of a microprocessor controlled PCB with sensor signal conditioning electronics, data storage, and RS-232 serial communication capability. This unit is internally powered by a 3.6 V lithium thionyl chloride cell battery, which provides monitoring lifetime of 5–7 years depending on duty cycle and temperature conditions. Prior to collecting corrosion data the AN101 was calibrated for API 5L ERW Grade-B μLPR sensors and a burn in was performed for a period of a month. During the experiment data was collected at 5 min intervals.

In addition to monitoring corrosion activity occurring on the pipe, data was collected concerning the moisture content, temperature, and salinity of the soil. This data was acquired at 5 min intervals with a 4-level soil moisture, temperature, and salinity probe placed along the vertical axis of the pipe. Data from the soil sensor was logged with an SDI-12 interface to a laptop computer and plotted in MATLAB.

Fig. 6 Analatom's AN101 corrosion monitoring node

3 Results and Discussion

Data collected and analyzed during the first month of dry and wet cycles indicated no significant corrosion had taken place on the sensors under the protective fusion-bonded epoxy coating of the mock natural gas pipeline. However, sensor 4, which was placed on top of the protective coating and exposed to the soil indicated within the first few days of the experiment, showed a modest amount of corrosion. This rate accelerated during the first wetting cycle which was initiated on day 7. As this sensor was directly exposed to the corrosive environment a fast response was anticipated. The total surface loss on sensor 4, the bare sensor, in Fig. 7 is approximately 25 μm or 10 % of the surface. At this point the interdigitated electrodes began to bridge, causing future LPR data collected from this sensor to become unreliable. This figure demonstrates the effect moisture content has on the corrosion rate.

After day 17, there was still no activity detected on sensors 1–3 and 5–8. During the start of the third wetting cycle on day 35, experimental parameters were modified to accelerate the corrosion process. To increase the rate of corrosion it was necessary to increase the temperature of the environmental chamber, which increased the diffusion coefficient and reduced the activation energy required for

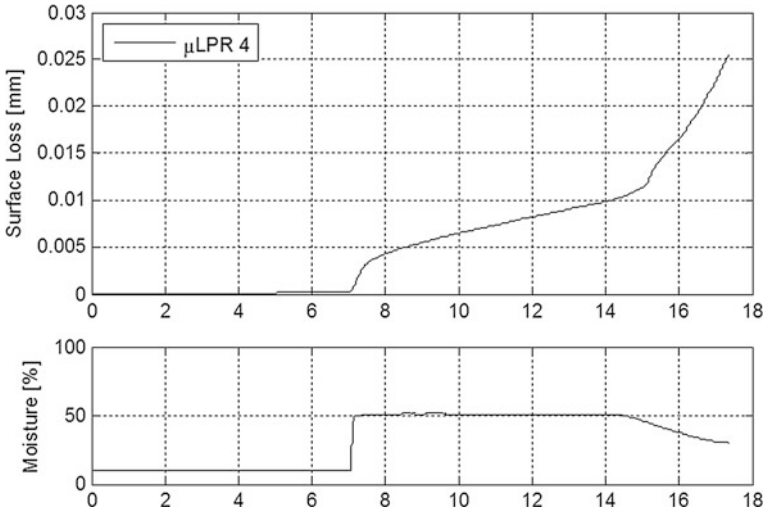


Fig. 7 Bare sensor surface loss and moisture during the first 17 days

electrochemical reactions to take place. To initiate this process the soil was held in a wet cycle with the infrared heat lamp system on, which had previously only been used to dry the soil. After a period of 14 days the average soil temperature began to approach 40 °C. At this point sensors five, six, and eight began to show LPR activity. All of these sensors were positioned along the pipe seam and were 18-in, 4 1/8-in, and 3/8-in away from the seam defect, respectively. All environmental and surface loss data collected is shown in Fig. 8.

Analysis of the period over which corrosion has been occurring for sensors under fusion-bonded epoxy provides more quantitative information pertaining to the surface loss of sensors. These data show over the 12 day period plotted in Fig. 9 that sensor six, the sensor 4.125-in away from the seam defect loses 7.0×10^{-2} μm or 0.028 % of its surface. Sensor eight, located 0.375-in from the seam defect lost approximately 3.9×10^{-2} μm or 0.015 % of its surface. Sensor seven, located 1.25-in from the seam defect lost approximately 1.2×10^{-2} μm or 0.005 % of its surface. Finally, these data show activity as far as 18-inches from the defect site, where sensor five lost 0.6×10^{-2} μm or 0.002 % of its surface.

Closer analysis of the sensors adjacent to the weld defect indicates these sensors are also experiencing surface loss. These data show over the 12 day period that sensor one, which is located 0.375-in from the weld defect lost 0.88×10^{-6} mm of its surface. Additionally, these data indicate corrosion activity is beginning on sensor two, which is located 1.25-in from the weld defect. A plot of the sensors adjacent to the weld is shown in Fig. 10.

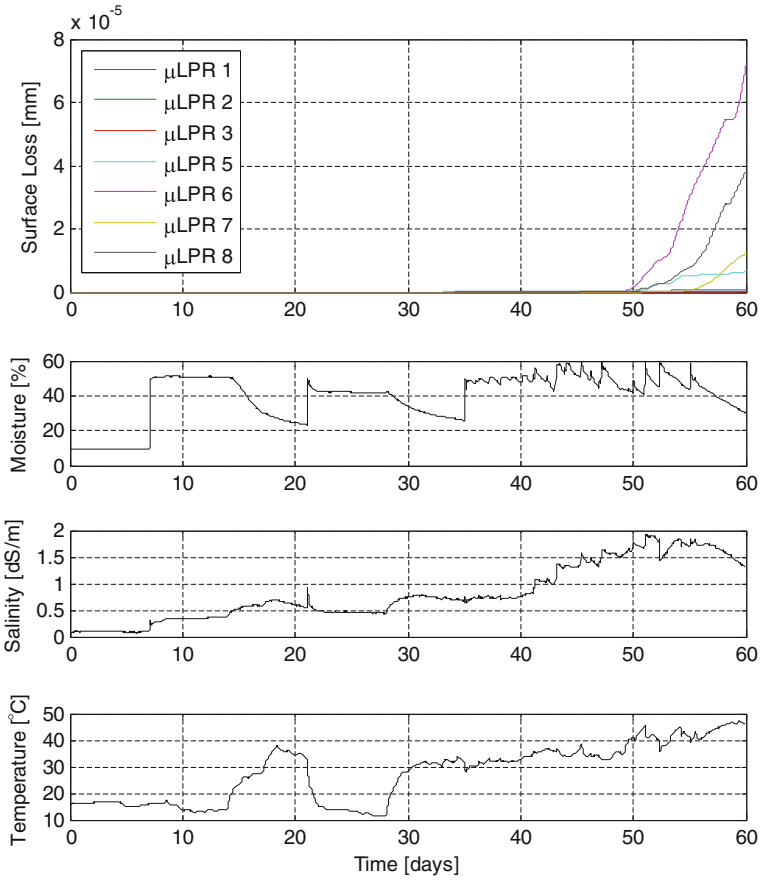


Fig. 8 Environmental and surface loss data for sensors under fusion-bonded epoxy

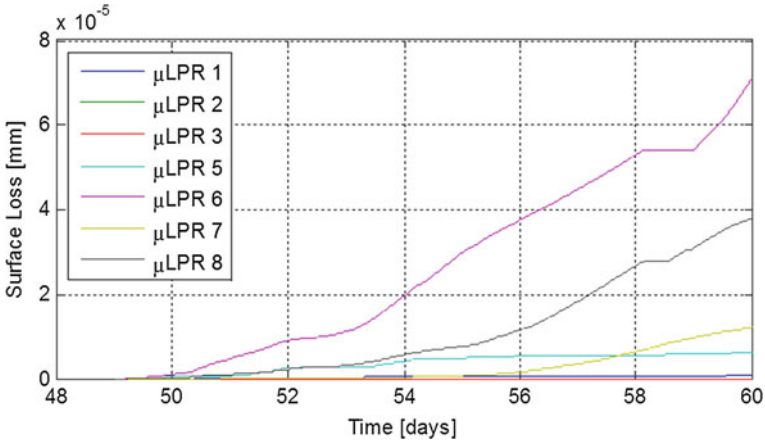


Fig. 9 Surface loss data for sensors under fusion-bonded epoxy from day 48 to 60

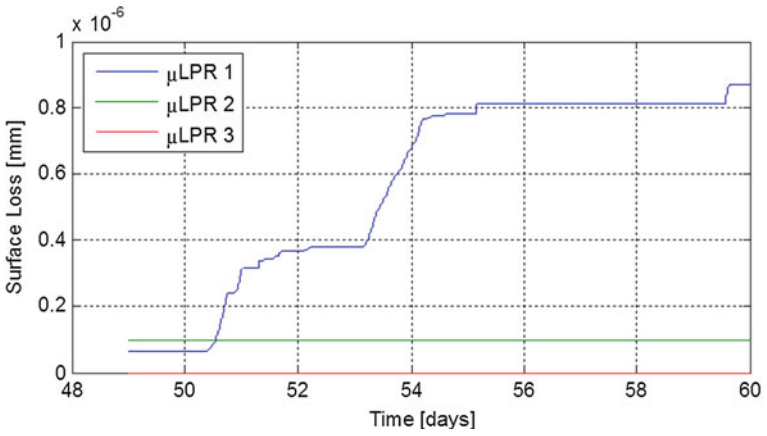


Fig. 10 Surface loss data for adjacent to weld defect from day 48 to 60

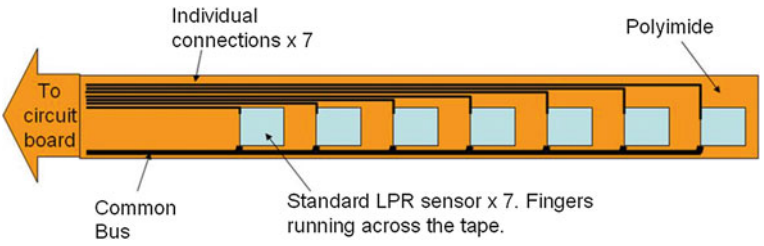


Fig. 11 Flexible tape design for μLPR sensors

4 Conclusions

Data gathered from sensors over the course of this experiment indicates pipe seams are highly susceptible to rapid infiltration of corrosive species. This is evident by all sensors along the seam showing surface loss, even as far as 18-in away from the defect. Oxidation occurring over such a large distance indicates undercutting of fusion-bonded epoxy occurs more readily along the seam, allowing species to rapidly diffuse along this boundary. Additionally, LPR activity observed on the sensor adjacent to the weld joint defect supports these conclusions, as the surface loss rate is orders of magnitude smaller than along the seam.

This experiment also indicates Analatom's μLPR sensors can be installed on natural gas pipelines without compromising the integrity of the surrounding fusion-bonded epoxy coating. While removal of the fusion bonded epoxy was necessary for sensor placement, the 3 M Scotchkote Liquid Epoxy 323 served as a viable alternative and prevented undercutting where wiring exited the coating. Failure of this coating material would have been evident by all sensors indicating surface loss from direct diffusion of corrosive species onto LPR sensors. These conclusions are

critical as one application of this technology involves placing μLPR sensors on pipes that are excavated for repair. This application would enable remote monitoring of the repair site and would serve as a critical indicator in corrosion hotspots, where soil properties are known to accelerate degradation of protective coatings. An additional application of this technology involves placing sensors in areas where pipes commonly corrode, such as the seam and near weld joints, prior to coating with fusion-bonded epoxy. To accomplish this a polyimide backing material with a higher melting point has been identified that would survive the high temperatures used in the application of fusion-bonded epoxy. Sensors installed in these hotspots during pipe fabrication will minimize the number of coating intrusion points by encompassing multiple sensors on one flexible circuit, shown in Fig. 11.

References

1. Sørensena PA, Dam-Johansena K, Weinellb CE, Kiil S (2010) Cathodic delamination: Quantification of ionic transport rates along coating–steel interfaces. Prog Org Coat 68:70–78
2. Pommersheim J, Nguyen T, Zhang Z, Lin C (1995) Cation diffusion at the polymer coating/metal interface. J Adhes Sci Technol 9(7):935–951
3. Cavassi P, Cornago M (1999) The cost of corrosion in the oil and gas industry. J Protective Coat Linings 16(5):30–40
4. CC Technologies Laboratories, Inc. (2001) Corrosion Costs and Preventive Strategies in the United States. Office of Infrastructure Research and Development, Federal Highway Administration
5. Huston D (2010) Structural sensing, health monitoring, and performance evaluation. Taylor and Francis, London
6. Scully JR (2000) Polarization resistance method for determination of instantaneous corrosion rates. Corrosion 56(2):199–218

Experimental Research on Diagnosis of Valve Leakage for Diesel Engines Based on Acoustic Emission

Yonghua Yu, Pengfei Ji and Jianguo Yang

Abstract The fault diagnosis mechanism of valve leakage for diesel engines based on acoustic emission signal is analyzed through a series of static and dynamic experiments in this work. Charging the cylinder with compressed air when the diesel engine is at the standby state, acoustic emission signals under normal and different degrees of valve leakage conditions were tested, characteristic frequency stimulated by valve leakage was found out. Then dynamic tests on diesel engine were further carried out, and the sensitive fault characteristic parameters were extracted. Finally, a method of identifying fault of valve leakage based on SVM model was presented according to these features.

Keywords Diesel engine · Acoustic emission · Valve leakage · SVM

1 Introduction

The cylinder valves are probably the most critical parts of diesel engines, because they must open and close automatically at every cycle, but they are subject to corrosion and wear due to their severe working condition of high temperature gas and combustion products as well [1]. As a result, the leakage of exhaust valve is likely to take place, which will have a very serious influence on the performance of diesel engines.

Acoustic emission (AE) measurement techniques can be used to detect internal and external leaks of process plant equipment like valves, steam traps, pipelines,

Y. Yu (✉) · P. Ji · J. Yang
School of Energy and Power Engineering, Wuhan University of Technology, Wuhan 430063, People's Republic of China
e-mail: yyhua@whut.edu.cn

Y. Yu · J. Yang
Key Laboratry of Marine Power Enigeering Granded by Ministry of Communication, 430063 Wuhan, People's Republic of China

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_59

and tanks. Compared with other means of detection, AE signal contains more information about the leakage sources due to its wide operating frequency and high signal-to-noise ratio. It has been mentioned by a number of authors that AE method is one of the most effective tools for detecting the ultrasonic high frequency emissions generated by a fluid leaks through a restricted orifice [2, 3].

The vibration monitoring and diagnosis research shows that the ultrasonic generated by valve leakage will produce a specific band width frequency in the cylinder head, which is always higher than the frequency produced by the explosion pressure. Therefore, valve leakage will force the vibration energy transfer to higher frequency band [4]. It is apparent that AE signals have a similar or the same characteristics once the leakage occurs. In order to investigate the diagnosis mechanism of valve leakage based on AE signals, a series of static and dynamic test scheme are put forward in this work.

2 Static Experiment

2.1 Experimental Apparatus

Experiments were carried out on the cylinder head of a 44.1 kW four-stroke 4120SD1 diesel engine. In the state of engine shutdown, valve stems and the rocker arm were demounted, and the piston was regulated to the BDC position before charging the cylinder #1 with 25 bar compressed air. Then AE signals under five different valve conditions were tested. All those work were done to make sure that the leakage AE source was isolated from other interference sources.

Figure 1 shows the data acquisition (DAQ) system used for the static experiment. A Physical Acoustic Corporation (PAC) Micro-80D AE sensor was coupled to the cylinder 1# surface by means of magnetic hold-downs(cylinder 1 and 2#

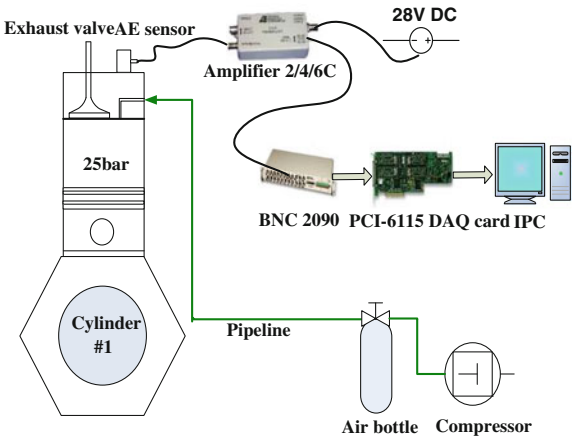


Fig. 1 Experimental apparatus

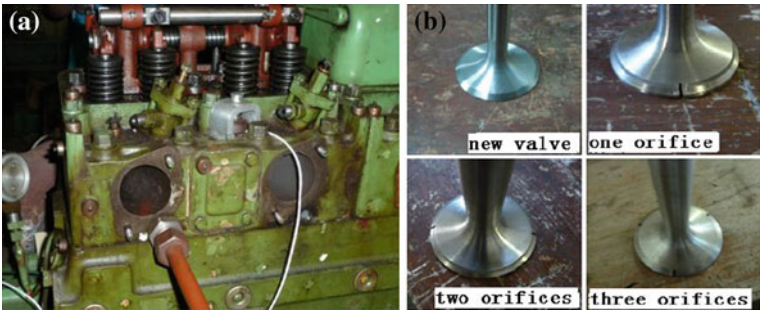


Fig. 2 **a** AE sensor installation. **b** Four kinds of leakage valves

share one cylinder head). The 4-channel National Instruments (NI) PCI-6115 DAQ card installed in an industrial personal computer (IPC) was used to acquire raw data with a sampling frequency of 2 MHz. AE signals under normal valve and four different conditions of leakage valves (Fig. 2b) were tested when the air bottle was opened.

2.2 Results and Discussion

A typical raw AE signal obtained from the experiment such as two orifices valve can be seen in Fig. 3. The time domain signal can be decomposed into three parts: the first part, signal before aeration; the second one, signal during aeration during the cylinder pressure was increased gradually to balance with the air bottle; the third one, signal after aeration while the cylinder pressure was equal to 25 bar.

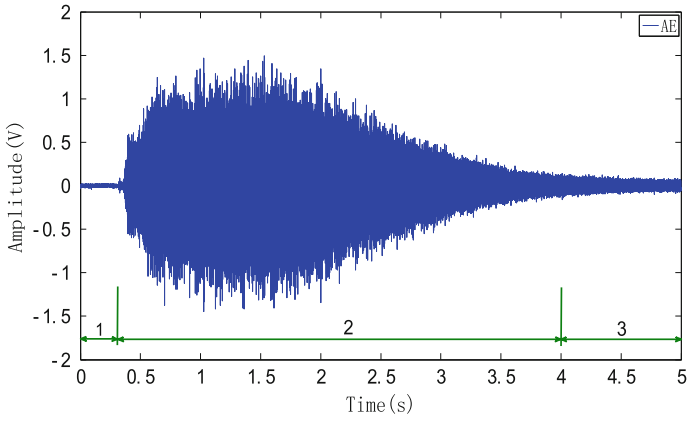


Fig. 3 A raw AE signal of a valve with two orifices

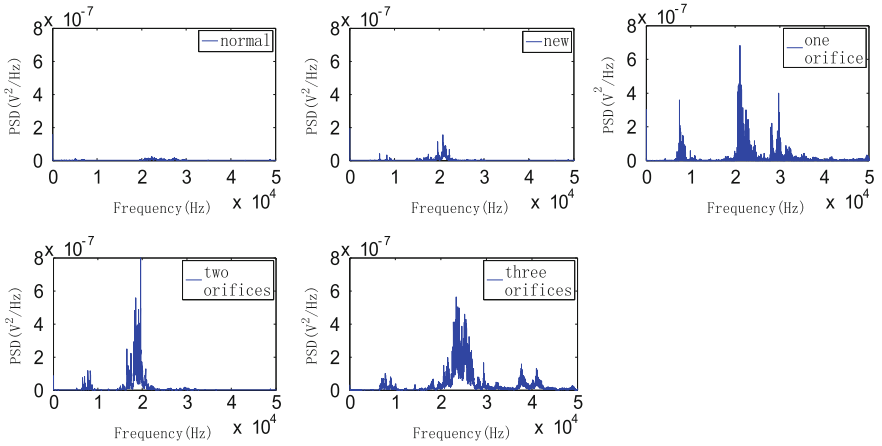


Fig. 4 PSD of AE signal under different valve conditions at static state

The impact generated by leakage will results in a typical AE source, which is constant at the third part of the whole signal. Therefore, the third part was focused to analyzed and processed.

Figure 4 shows PSD of the third part of AE signal under different valve conditions. It is concluded that:

1. The frequency of AE signal generated by valve leakage is mainly within the range of 20 $\sim$ 30 kHz and associated with the form and the degree of leakage.
2. In the state of normal valve, the AE signal energy is much small because there is theoretically no leakage at all.
3. The new valve without grinding does not show strong incentive frequency of leakage, not only because of its mild leakage form but also the limited pressure of static experiment.
4. The rest of three valves showed strong incentive frequency of leakage, but the main frequency distribution is slightly different, the frequency of one orifice valve mainly concentrated in the 20 $\sim$ 30 kHz as well as three orifices valve, however the excitation frequency of two orifices valve was mainly within the scope of 15 $\sim$ 20 kHz.

3 Test on Running Engine

To gain a greater understanding of valve leakage events it is necessary to further investigate the feature of the AE signals. A dynamic test research on running engine is indispensable. In this section, a dynamic test in various valve states was carried out on a running 4120SD1 diesel engine.

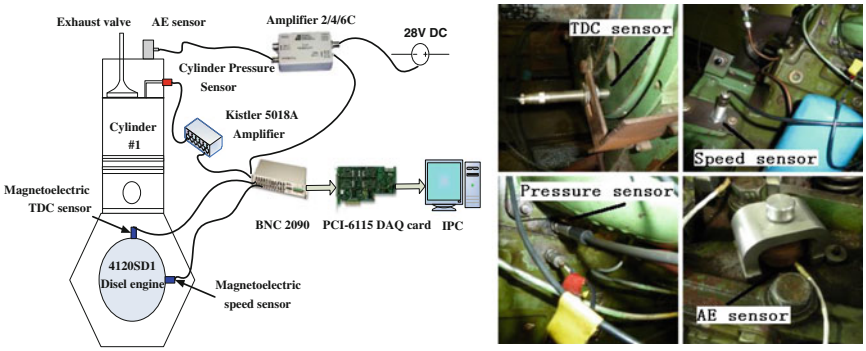


Fig. 5 **a** Test apparatus **b** Sensor positions

3.1 Experimental Scheme

A Micro-80D AE sensor, a Kistler 7013C cylinder pressure sensor and two magnetoelectric sensors were installed on the 4120SD1 diesel engine. The NI PCI-6115 DAQ card was used to acquire raw data with sampling frequency of 2 MHz, as shown in Fig. 5. AE signals under normal valve and four different conditions of leakage valves (same as Fig. 2b) were tested at load characteristic curve, five load conditions ranging from 0 to 90 % at 1,500 r/min.

3.2 Power Spectral Analysis

For a running engine, it is expected that the acquired signal will be a combination of mechanical impacts and fluid excitation coming from injectors, intake valves Open/Close (IVO/C), Exhaust Valves Open/Close (EVO/C), combustion and also ancillary equipment. The TDC signal and cylinder pressure signal were used as a trigger to start acquiring data from the AE sensor and the timing signal. The valve timing of the diesel engine can be used to identify all those exciting sources, as shown in Fig. 6. A typical raw AE signal in a cycle on the cylinder head can be seen in Fig. 7.

The valve leakage mostly influences AE signals in combustion process when all valves are closed and the pressure inside the cylinder is particularly high. Thus a part of the AE signal from 347 to 380 °CA was selected to calculate PSD in different valve and load conditions, as shown in Fig. 8.

It can be concluded that:

1. The frequency generated by explosion pressure in normal valve state was mainly within the scope of 18.5 ∼ 22 kHz and less affected by different load conditions.

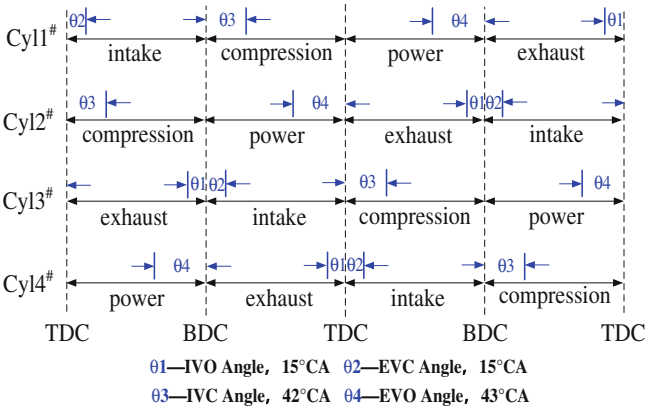


Fig. 6 Valve timing of the 4120SD1 diesel engine

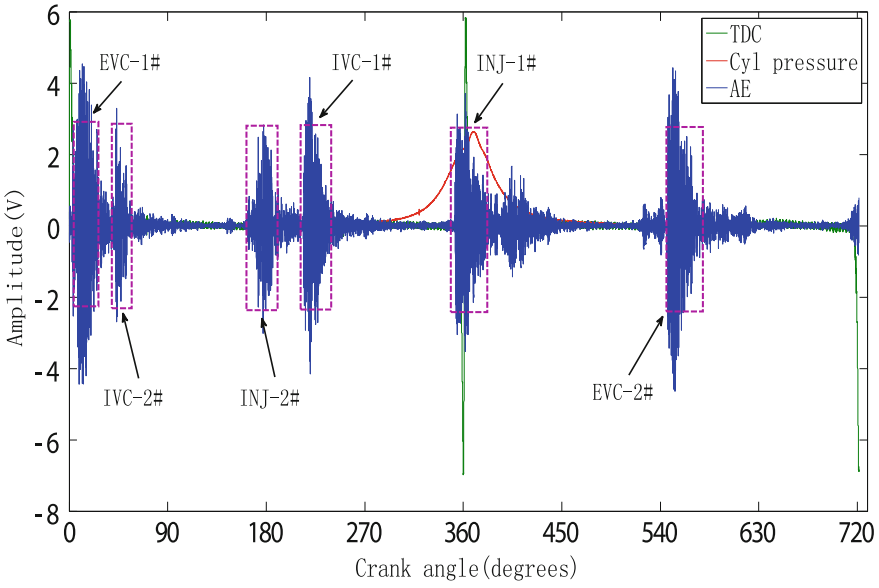


Fig. 7 Raw signals in a cycle

2. Variations of PSD will appear when valves leaking. The energy of AE signal transfers to a higher 22 ~ 30 kHz frequency band under valves leakage conditions.
3. The results of frequency-domain analysis at running test are consistent with those of static experiment.

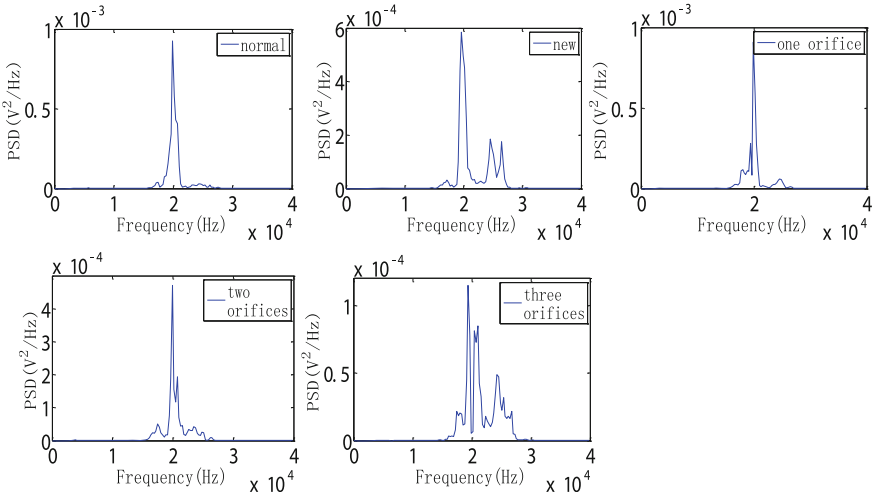


Fig. 8 PSD of AE signal under different valve conditions at 1,500 r/min, 90 % load

3.3 Feature Extraction

Two sensitive AE characteristic parameters P1/P and PSD2/PSD1 were extracted according to the previous analysis, as shown in Fig. 9.

As shown in Table 1, P1/P become less, while PSD2/ PSD1 become larger than that of normal condition when valve leak. An optimal criterion can be established to detect the valve leakage condition using these two parameters like these: (1) P1/P > 85 %; (2) PSD2/PSD1 < 10 %.

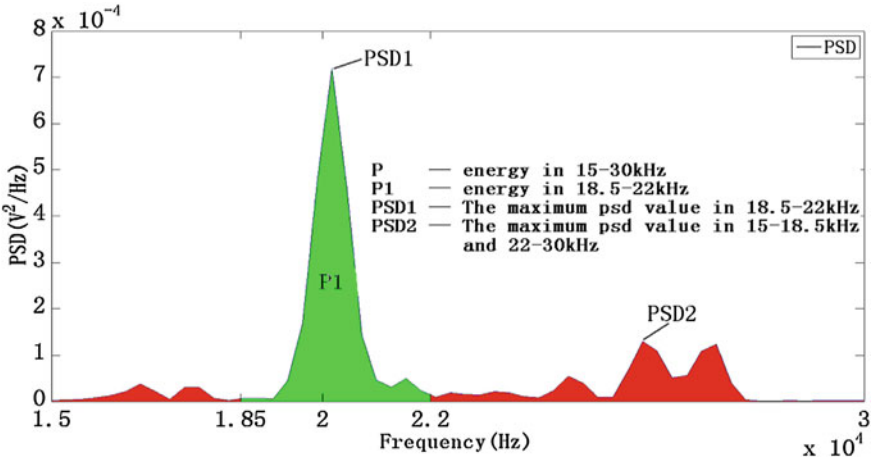


Fig. 9 Definition of AE characteristic parameters

Table 1 Values of AE characteristic parameters under different valve and load conditions

Load (%)	Parameter	Valve conditions				
		Normal (%)	New (%)	One orifice (%)	Two orifices (%)	Three orifices (%)
0	P1/P	89.2	66	83.9	73	71.9
	PSD2/PSD1	4.9	18	10.4	20.3	24.2
25	P1/P	90.4	74.6	65.7	69.4	73.3
	PSD2/PSD1	5.6	14.4	19.4	24.2	18.4
50	P1/P	92.5	62.3	69.5	74.6	73.9
	PSD2/PSD1	2.5	24.3	19.7	10.9	16.3
75	P1/P	87.9	61.1	83.8	75.4	63.7
	PSD2/PSD1	4.8	23.4	7.6	12.9	49.8
90	P1/P	87.2	62.6	74.6	70.4	53.5
	PSD2/PSD1	4.2	31.6	12.8	10.8	42.6

4 Fault Recognition Based on SVM

4.1 Feature Extraction

Two frequency domain parameters P1/P and PSD2/PSD1 were extracted in Sect. 3. 3 cyclic parameters and 1 maximum PSD are added in order to improve the classification accuracy. A Support Vector Machine (SVM) model is used to classify the experimental AE data as shown in Table 2.

4.2 Results

There are 250 testing samples under normal and different valve leakage at each work condition, 150 samples are chosen as the training samples of SVM model, that is, 30 samples for each valve condition. Target samples of normal valve, new valve, one orifice valve, two orifices valve and three orifices valve are labeled 0,1,2,3,4 respectively. After the SVM model has been trained by the 150 samples, it has capability of identifying valve leakage. Another 100 testing samples are chosen to

Table 2 AE feature vectors for SVM model

Vector	Definition
Rms1	Rms value of AE signal in combustion period (332 ~ 400 °CA)
Rms2	Rms value of AE signal in EVC (0 ~ 30 °CA)
Rms3	Rms value of AE signal in a cycle (0 ~ 720 °CA)
P1/P	The ratio of AE enerey in 18.5 ~ 22 and 15 ~ 30 kHz
PSD2/PSD1	The ratio of the max PSD value in 15 ~ 18.5, 22 ~ 30 and 15 ~ 30 kHz
PSDmax	The max PSD value in 15 ~ 30 kHz

Table 3 Testing samples of eigenvector at 1,500 r/min, 90 % load

No.	Actual state	Label	Feature vector	Identifying result
1	Normal	0	(1.017 0.978 0.453 0.889 0.036 0.402)	Normal
2	Normal	0	(0.980 0.959 0.450 0.868 0.046 0.454)	Normal
3	New	1	(1.055 1.245 0.494 0.673 0.351 0.209)	New
4	New	1	(1.069 1.257 0.486 0.731 0.134 0.392)	New
5	One orifice	2	(1.071 0.874 0.507 0.861 0.116 0.329)	One orifice
6*	One orifice	2	(1.028 0.828 0.503 0.851 0.140 0.249)	Normal*
7	Two orifices	3	(0.811 0.762 0.371 0.785 0.217 0.115)	Two orifices
8	Two orifices	3	(0.842 0.744 0.400 0.786 0.201 0.153)	Two orifices
9	Three orifices	4	(0.677 0.450 0.393 0.660 0.305 0.067)	Three orifices
10	Three orifices	4	(0.668 0.426 0.380 0.655 0.233 0.064)	Three orifices

Table 4 Classification accuracy

Valve State	Load (%)				
	0 (%)	25 (%)	50 (%)	75 (%)	90 (%)
Normal	95	90	95	95	90
New	85	80	75	80	100
One orifice	100	100	100	80	75
Two orifices	95	100	90	100	90
Three orifices	80	75	100	100	100
Average	92	89	92	91	91

verify the SVM model. Some testing results in load 90 % are illustrated in Table 3 as an example (* represents a wrong result in Table 3). The whole classification accuracy under different valve and load condition are listed in Table 4.

It is shown that most testing samples are identified well except in some isolated cases such as new valve in 50 % load and one orifice valve in 90 % load. As a result, the average rate of successfully identifying is about 90 %. It means that the model designed possesses the capability of identifying valve leakage. There is actually no need to identify so many valve leakage conditions, therefore, the classification accuracy of this SVM model is accurate enough for detecting whether a valve has leakage or not.

5 Conclusions

1. As a specific AE source, valve leakage stimulates a particular frequency bandwidth in the cylinder head which associated much with the form and the degree of leakage.

2. Experiment test on running 4120SD1 diesel engine shows that the plot of PSD in combustion period changes distinctly when a leakage take place, two frequency-domain parameters P1/P and PSD2/PSD1 can be used as criterion to detect whether there is any leakage.
3. It is feasible to use SVM classification technique to achieve intelligent diagnosis for small AE samples. The testing result shows that the SVM model established by six-dimensional AE feature vectors which consist of Rms1, Rms2, Rms3, P1/P, PSD2/PSD1 and PSDmax is effective to identify valve leakage mode.

References

1. Dickey J, Dimmick J, Moore PM (1978) Acoustic measurement of valve leakage rates. Mater Eval 36:67–77
2. Püttmer A (2007) Acoustic emission based online valve leak detection and testing. IEEE Ultrason Symp 269:1854–1857
3. Pornchai N (2004) Multi-source, multi-sensor approaches to disel engine monitoring using acoustic emission[D], Heriot-Watt University, Edinburgh
4. Jianguo Y (1997) Research on diagnosis of valve leakage for diesel engines based on vibration. Mar Technol 19(6):41–44

Development of Safety, Control and Monitoring System for Medium-Speed Marine Diesel Engine

Qinpeng Wang, Yihang Qin, Jianguo Yang, Yonghua Yu and Yuhai He

Abstract A local safety, control and monitoring system designed and implemented for medium-speed marine diesel engine is presented in the paper. The system consists of a control and monitoring sub-system, a safety sub-system and an online monitoring sub-system. The data exchange among the sub-system is conducted with the communication link of RS-485 interface. The hardware circuits and logical algorithm of the system are developed to achieve the functionalities in terms of diesel engine staring, stopping, fault shutdown protection, etc. The operating data of the diesel engine are captured in real time and the vital operational parameters and alarm messages are displayed with the nixie tubes and LED. As well, depending on the online monitoring sub-system, the control instruction and operational parameters could be monitored, analyzed and recorded. Additionally, a test platform is established to verify and validate the function of the system. Finally, the matching test between the system and the diesel engine is carried out via the experiment bench of MAN16/24 medium-speed diesel engine. The results show that the local safety, control and monitoring system works stable and the functions of the conventional control, parameter monitoring and alarming for the diesel engine meet the design requirement.

Keywords Control and monitoring system · Safety · Marine diesel engine

Q. Wang (✉) · Y. Qin · J. Yang · Y. Yu · Y. He
School of Energy and Power Engineering, Wuhan University of Technology, Wuhan 430063, P.R.China
e-mail: wangqpkevin@163.com

J. Yang
e-mail: jgyang@whut.edu.cn

J. Yang
Key Laboratory of Marine Power Engineering &Technology Under Ministry of Communications, Wuhan 430063, P.R.China

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_60

1 Introduction

Intelligence is the development tendency of the marine automation control technology, based on the rapid technological development and the reduction of ship's operational costs. The digitization, network, and modular are the points of the control system for maritime applications [1–3]. A local control and monitoring system is the significant component of the control systems of marine diesel engines. The paper reports on the design and implement of a local safety, control and monitoring system for medium-speed marine. Depending on a modular system philosophy, the system is composed of control and monitoring sub-system, safety sub-system and online monitoring sub-system. The functionalities are achieved including the engine staring, stopping, operational parameters display, alarms and security protections.

2 Structure and Function of the System

The local safety, control and monitoring system is an integrated control system with the functions of monitoring, control, protection and alarm. A modular concept of components which can be applied flexible according to different functions, builds the basis. An identical communication is available for data exchange among functional modules, and at the hardware and software levels each modules are separated. The structure of the entire system is illustrated on Fig. 1.

3 Control and Monitoring Sub-system

The control and monitoring sub-system is the core part, and is consisted of master control module, operation module and display module.

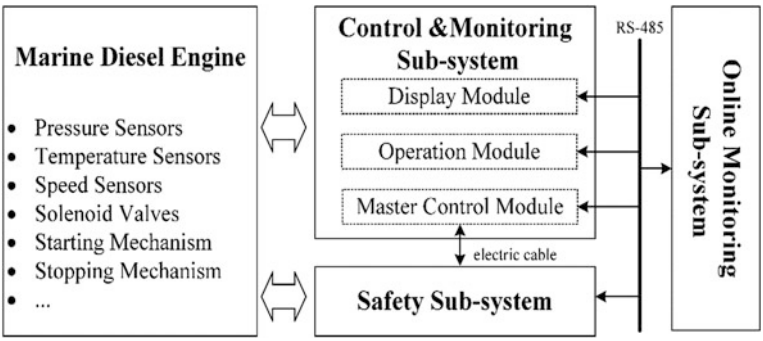


Fig. 1 System architecture

3.1 Master Module

STM 32 microcontroller chip of ST Corporation based on the ARM Contex-M3 kernel is adopted for the master module [4]. Specific peripheral circuits are designed for the functionalities of signals acquirement, driving and communication. The circuits mainly include various measure circuits, signal modulating circuit, isolating driver circuit, communication circle and electrical source.

Signals from sensors assembled on the diesel engine are collected by the master module in real-time for monitoring the operating states of diesel engines. Additionally, some of the operating parameters and warming signals are transferred to display module via RS-485 bus. The acquiring signals contain the revolution speeds of the diesel engine and the turbocharger, lubricating oil temperature, exhaust gas temperature, fuel pressure and so on. The conversion from sensor signals to physical quantities is calculated as follows.

1. Rotational speed

Depending on the crankshaft rotation, an approximate sinusoidal signal is generated from the speed sensor, and the signal is converted to pulse signal by means of the speed measurement circuit. The rotational speed is deduced in the Eq. (1).

$$n = \frac{60C}{Z * T} \tag{1}$$

where n is the diesel engine speed; C is the plus number in the count cycle; T is the count cycle; Z is the Tooth plate number.

2. Temperature calculation

Temperature calculation is mainly used for resistance temperature and thermocouple sensors. Temperature is taken as a function of the voltage signal by indexing the reference tables of resistance and thermocouple sensors, and temperature values are obtained with the linear interpolation method.

$$T_{pt} = f(\Omega) \tag{2}$$

$$T_{tc} = f(v_{tc}) \tag{3}$$

where T_{pt} and Ω are the value and signal of the resistance temperature sensor; T_{tc} and v_{tc} are the value and signal of the thermocouple sensors.

3. Pressure calculation

The output currents are the standard 4–20 mA, and the current signal has a linear proportional relationship with the pressure value. The physical pressure value can be expressed as:

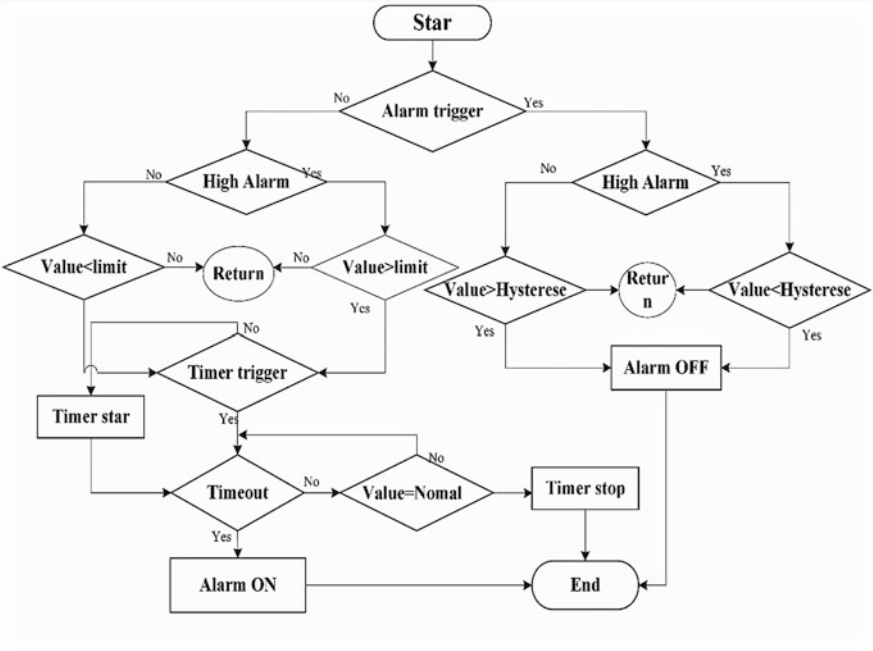


Fig. 2 Alarm logical flow chart

$$P = I\frac{H_{\max} - H_{\min}}{20 - 4} \tag{4}$$

where P is the pressure value; $H_{\max}$ the maximum rang of the sensor; $H_{\min}$ is the minimum range of the sensor; I is the output current signal.

The logical function for judging operating parameters out-of-limit, is designed in the master module. When some parameter is exceeded the normal range, the annunciator will be active according to the hysteresis principle. The logical flow is presented on Fig. 2.

In accordance with the commands from the operation module, the control of the starting electromagnetic valve, the stop solenoid valve, the fuel limit valve and others is achieved by the master module. The control flow is illustrated on Fig. 3.

3.2 Operation and Display Module

The operation module and the display module are the human-computer interaction devices, and the data exchange with the master control module.is through the RS-485 bus. The commands such as starting, stopping, remote/local switch and alarm reset, are sent out from the operation module. Besides, all the operating

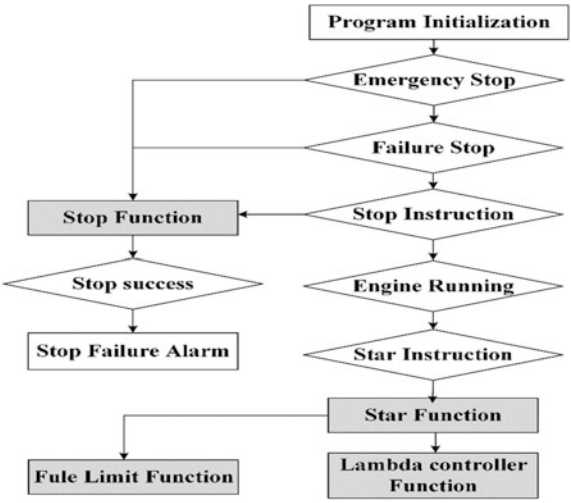


Fig. 3 Control function flow chart

parameters collected by the master module can be shown in the liquid crystal display. The display module is mainly composed of LED lights, and is used for displaying the essential temperature and pressure parameters. Simultaneously, the off-limit alarm and disconnection alarm of sensors are also in effect.

4 Safety Sub-system

The safety sub-system is a dependent protection system of diesel engines for emergency stop. It is regarded as a secondary protection device due to lower thresholds. The programmable logic controller (PLC) is selected as the control core and is dedicated to data acquisition. The function of the sub-system is to identity the emergency situation of the diesel engine and stop the diesel engine from crashing. The diesel engines will be enforce to stop in the following conditions:

(1) Over speed.
(2) Low pressure of lubricating oil.
(3) High temperature of cooling water.
(4) Emergency stopping.

5 Online Monitoring Sub-system

Based on the LabVIEW software, an online monitoring sub-system is developed for monitoring the instructions and the operating data from other sub-systems via RS-485 bus, and it has the abilities of data analysis and files recording. The overview of the software is described on Fig. 4.

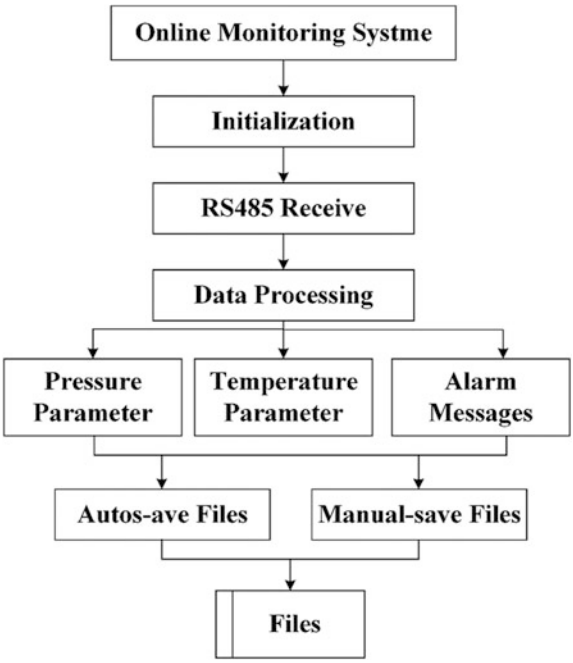


Fig. 4 Program architecture

6 Development of Test System

For the functional validation of the whole system, a test system is established depending on the NI hardware and software platform. The hardware consists in a NI-PXI 8176 controller, a NI-PXI 7853 data-acquisition card and C-series I/O modules. In terms of Mean-value model principle [5], a diesel engine is modeled with the experimental data maps. The key working parameters of a diesel engine are simulated for providing the necessary boundaries of the whole system. The construction of the test system is introduced on Fig. 5.

7 Matching Test

The functionalities of the developed system are validated with the help of an experiment bench of a MAN.B&W L16/24 diesel engine. The control and monitoring sub-system, and the safety sub-system are integrated in a box installed on the diesel engine. The scene is shown on Fig. 6.

The testing items cover the function of emergency stop, starting interlock, fault shutdown, restarting, failure stop and so forth. In addition, the load characteristic testing of the diesel engine is accomplished with the typical operating points, and the experimental data are filed by the online monitoring sub-system.

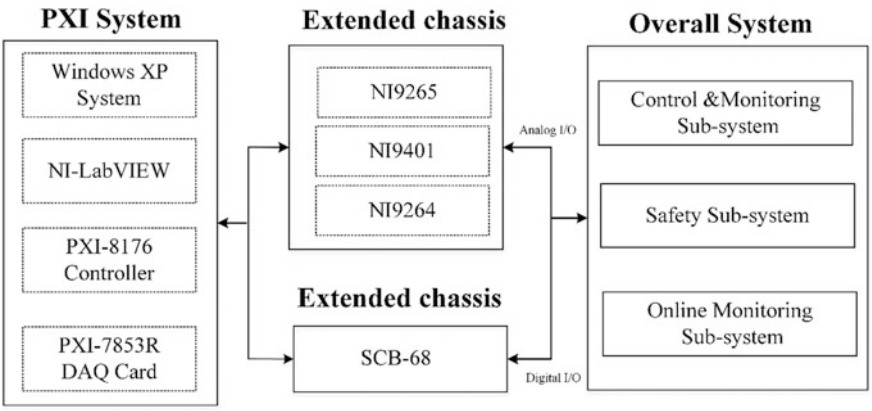


Fig. 5 Structure of the test platform

Fig. 6 Testing site

8 Conclusion

The local safety, control and monitoring system is designed and implemented. The system is composed of control and monitoring sub-system, safety sub-system and online monitoring sub-system. RS-485 bus is used for data exchange among the sub-systems. The function of the routine control, parameters monitoring and alarming is implemented. The matching experiment is completed with the experiment bench of MAN16/24 medium-speed diesel engine. The results show that the system works stable and is available for the engineering application.

References

1. Whittington HW, Jordan JR, Paterson N et al (2008) Performance monitoring of diesel electricity generation. IEEE Proc Electr Power Appl 133(3):149–154
2. Bian G (2007) Study on a monitoring and controlling system for the intelligentized de tection of ship's diesel engine. Marine Electr Electron Eng 27(6):356–381
3. Hu Y etc (2012) Marine diesel engine propulsion system. Shanghai jiao-tong university press, Shanghai
4. Meng B (2012). STM32 Notes. BEIHANG university press
5. Hendricks E (1989) Mean value modeling of large turbo charged two-stroke diesel engines. SAE Paper890564

Research on Magnetism Monitoring Technology of Piston Ring Wear for Marine Diesel Engine

Jian-guo Yang and Qiao-ying Huang

Abstract As the key parts of marine diesel engine, the working state of the piston rings affects the performance of the marine diesel engine directly. Therefore, it is significant to research on the monitoring piston rings wear method for the marine diesel engine. The three-dimensional finite element calculation magneto-resistive model of piston rings wear was developed based on RTA52U marine diesel engine in the paper. There is a single corresponding relationship between the piston ring wear and magnetic field changes on the monitoring point by theoretical calculation results. The sensor used for monitoring piston rings wear and a sensor calibration equipment were developed. The piston rings wear monitoring sensor is developed and reliable through calibration test. The correction of the calculation magneto-resistive model is verified by experiments on board.

Keywords Marine diesel engine · Piston ring wear · Magneto-resistive sensor · Three-dimensional magnetic field simulation

1 Introduction

A large low-speed marine diesel engine is the power source of a ship and its working fault has been threatening the ship's safety. Due to the poor working condition of the diesel engine, there is a great possibility of diesel engine working fault [1, 2]. As the key parts of diesel engine, piston rings operation state affects the

J. Yang (✉) · Q. Huang
School of Energy and Power Engineering, Wuhan University of Technology, Wuhan 430063, China
e-mail: jgyang@whut.edu.cn

J. Yang
Key Laboratory of Marine Power Engineering & Technology under Minister of Communication, Wuhan 430063, People's Republic of China

Q. Huang
China Classification Society Jiangsu Branch, Nanjing 210011, China
e-mail: qiaoying625@163.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_61

performance of the diesel engine directly. However, it is difficult to monitor piston rings wear on-line accurately considering those sealing working environment. The diesel engine faults caused by the piston rings wear of diesel engine occupy a large proportion of the shipping faults. Therefore it is significant to research the monitoring of the marine diesel engine piston rings wear [3, 4].

The RTA52U type Marine diesel engine is treated as the research object in this paper, and the magnetic three-dimensional finite element calculation model is established. The corresponding relation between the piston ring wear and the characteristic value of the monitoring magnetic field is analyzed through the finite element calculation model. Besides a sensor for piston ring wear monitoring and its calibration device is developed, then the sensor is calibrated through calibration device. Then actual ship test is banded to confirm the accuracy of the 3D calculation model and the diagnostic of the monitoring technology.

2 Piston Ring Wears Monitored Principle of Magnetic Resistance Sensor

Piston ring wear monitored principle of magnetic resistance sensor is through monitoring the changes of magnetic field strength, then the monitoring and diagnosis of piston ring wear conditions is achieved. Monitoring mechanism of magnetic resistance sensor is analyzed in the section.

2.1 Monitoring Mechanism of Magnetic Resistance Sensor

Magnetic resistance sensor is Wheatstone bridge circuit composed of four same permalloy thin film, as shown in Fig. 1. When external magnetic field is acted on

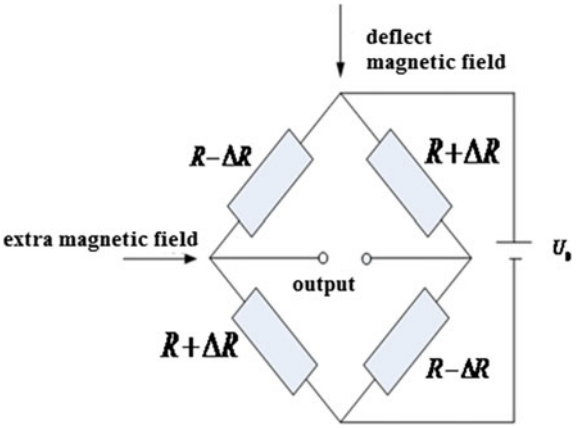


Fig. 1 Internal bridge of magnetic resistance sensor

the membrane, permalloy anisotropic magneto resistance effect occurs in the film, its internal resistance value changes, namely the bridge resistance in the circuit changes, the sensor output voltage value is changed, the output voltage equation is as follows [5, 6]

$$U_{out} = \frac{\Delta R}{R} V_b = kB_b V \tag{1}$$

with V_b as the Bridge work voltage, R as bridge arm resistance, $\Delta R/R$ as the relative rate of change of magnetic resistance, k as the sensitivity of sensors, B as the external magnetic induction intensity of magnetic field.

By formula (1), the sensor output voltage is proportional to the intensity of the external magnetic field, the magnetic resistance sensor output voltage is consistent with the external magnetic field changes, and so magnetic change can be monitored by magnetic resistance sensor.

2.2 Piston Ring Wears Electromagnetic Field Finite Element Analysis

This study takes 6RTA52U type diesel engine as the simulation object, the magnetic three-dimensional finite element calculation model of the 6RTA52U diesel engine is established with the finite element analysis software ANSOFT MAXWELL, as shown in Fig. 2, the model includes diesel engine piston, cylinder liner, piston ring, and magnetic sensors. The sensor consists of a permanent magnet, sensor chip and shell. Material properties of the model is set according to the characteristics of the diesel engine parts, the cylinder liner and piston set for iron material, piston ring and the sensor shell material as steel. The sensor chip monitors the change of the magnetic field intensity, and through the extraction of characteristic parameters as the basis of piston ring wear judgment. So the magnetic field intensity of the calculation point (monitoring) is taken as the output value of the magnetic sensor.

For the analysis of the effect of piston ring wear on output signal of the sensor, different wear condition of piston rings (different thickness) are generated into the magnetic three dimensional finite element model to calculate, using 3D static magnetic field solution parametric method the impact of the piston wear on the output signal is solved. Piston ring thickness is parameterized, by setting the amount change the thickness of the piston ring wear, in the process of calculation the external surface of the piston ring has always been close to the cylinder liner inner surface.

When piston ring wear, the amplitude of the magnetic field intensity of the monitoring point has a one-to-one correspondence relationship of piston ring wear quantity, magnetic field intensity amplitude declined with the increase of the piston ring wear. Because the magnetic resistance sensor monitoring of the magnetic field

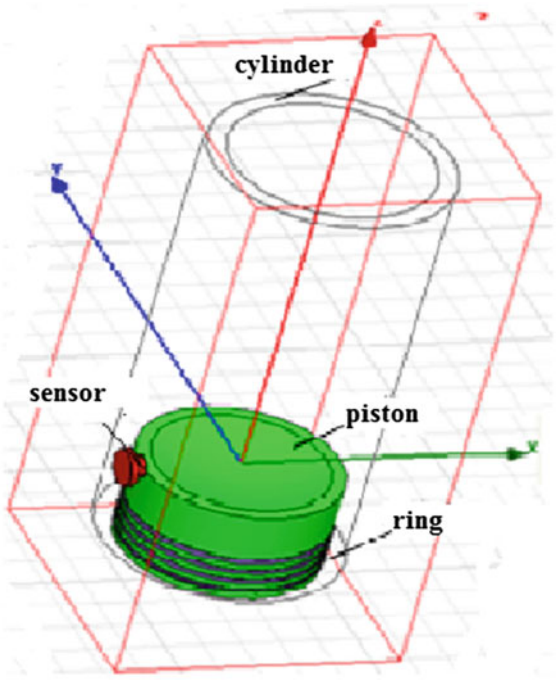


Fig. 2 The finite element calculation model

intensity, magnetic resistance sensor output voltage has a linear relation with the monitoring point magnetic field intensity, considering the sensitivity of the sensor, the simulation results is converted to sensor output voltage and the corresponding relation of piston ring wear as shown in Fig. 3. The piston ring wear and sensor output voltage amplitude is also one-to-one related, so the piston ring wear quantity can be monitored through the magnetic sensor.

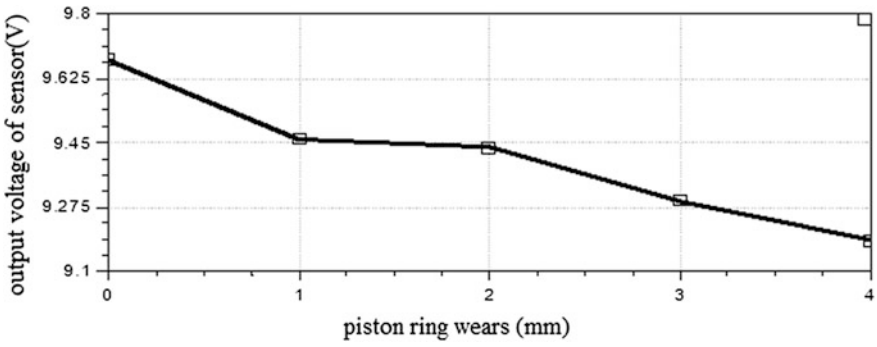


Fig. 3 Piston ring wear and the output of sensor

3 Development of Piston Ring Wear Monitoring Sensor

The monitoring piston ring wear mechanism based on magneto-resistive is to verity the corresponding relation between the degree of wear of the piston ring and the characteristic parameter of magnetic field strength at the monitoring point. Then the wear condition of piston ring can be judged by the different of characteristic parameters value. Under the same condition of external factors,change of characteristic value of magnetic field strength, which is caused by piston ring wear, is derived from magnetization of piston ring. The magnetization of piston ring is bought out by the permanent magnet inside the sensor. Magneto-resistive sensor chip monitors the magnetic field which is generated by the magnetized ring and judges the degree of wear of the piston ring [7, 8].

In determining the internal structure of the sensor, we take following into main consideration: choice of magneto-resistive sensor chip and permanent iron and determination of distance from permanent magnet to magneto resistive sensor chip. Consider the characteristics of wide range and high resolution of piston ring wear of actual marine diesel engine; we need to select magneto resistive sensor chip and permanent magnet according to the simulation result of piston ring wear and the requirements of range and sensitivity. The distance from magneto-resistive sensor chip to permanent magnet need adjustment in the actual test so that the magnetic field which generated by the permanent magnet ring can magnetize piston ring and in the scale range of the magneto-resistive sensor chip. The diagram of sensor is shown in Fig. 4. The Magneto-resistive sensor chip in the sensor is bridge circuit. In order to reduce magnetic interference from outside, the peripheral circuit of designed sensor is shown in Fig. 5.

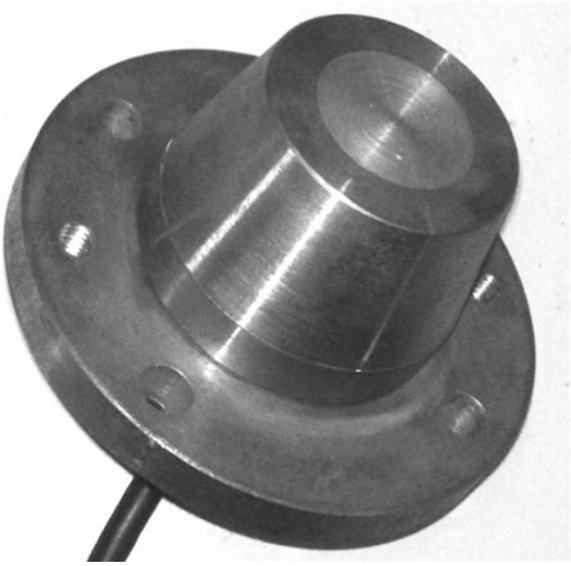


Fig. 4 Piston ring wears monitoring sensor

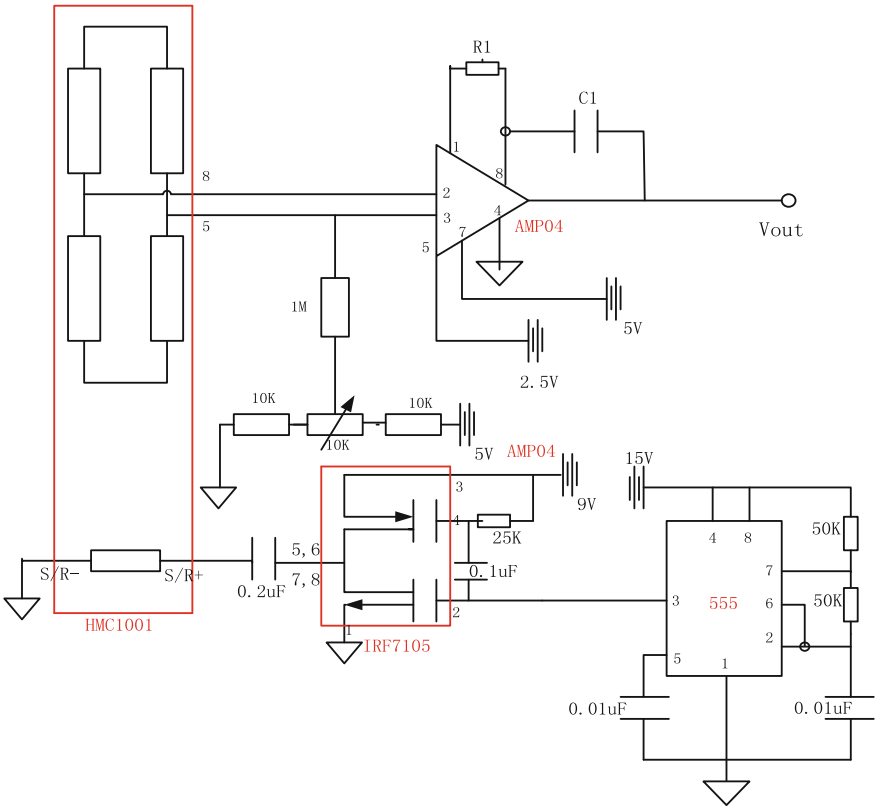


Fig. 5 External electro circuit of sensor

3.1 Monitoring Piston Ring Wear Sensor Calibration

The result of the FEM magnetism simulation model of piston ring wear is the relation between ring wear and the magnetic field, while the actual test output is voltage. Therefore, it is necessary to ensure the relationship between the output voltage and the magnetic field. It means the calibration of the measurement system.

The standard magnetic field which is generated through calibration device is provided to calibrate the piston ring wear monitoring system. The calibration device using he magnetic field from the position of the axis of the Helmholtz coils. There is relation between the magnetic field the coil current as follows [9]:

$$B_o = \frac{8}{5^{3/2}} \frac{\mu_0 \cdot N \cdot I}{R} \tag{2}$$

with R is the radius of the coil, N is coil turns, μ_0 is vacuum magnetic permeability, I is electric current.

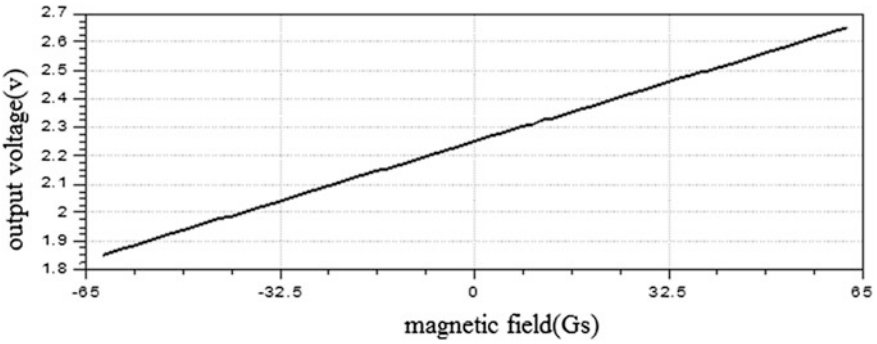


Fig. 6 The relation between output voltage and magnetic field

In the calibration test, sensor install position is asked to meet the reaction surface of sensor aligning with the central axis of the calibration device. The magnetic field at the axis of the calibration device is treated as known values. Output values of the sensor are recorded.

The correspondence between output voltages of the sensor and he magnetic fields is acquired through matching method, shown in Fig. 6. It means that the slope of the straight line is the sensor sensitivity, and the figure show that the sensitivity of sensor's is 6.402Gs/mV, linear range is (−62Gs–62Gs). It knows that the developed sensor has high sensitivity and wide linear range. It can meet a wide range of marine diesel engine piston ring wear and high resolution of technical requirements.

4 The Real Ship Test and the Verification of the Piston Ring Wear Monitoring

The real ship experimental is ask to verify correct and accuracy of calculation model. The piston ring wear of magnetic monitoring technology of marine diesel engine at the experimental point is studied in this section.

4.1 Measuring System and Ship Tests

A real ship trial is done in the Daqing 454 tanker, and the type of main engine is 6RTA52U. Fig. 7 shows the sensor actual installation diagram, it installed in the cylinder bore, fixed by bolts. Sensor install hole locate near the engine scavenging port, where the gas pressure is small and the temperature is relatively low. So there is little effect to the sealing of the combustion chamber and the sensor. The sensor

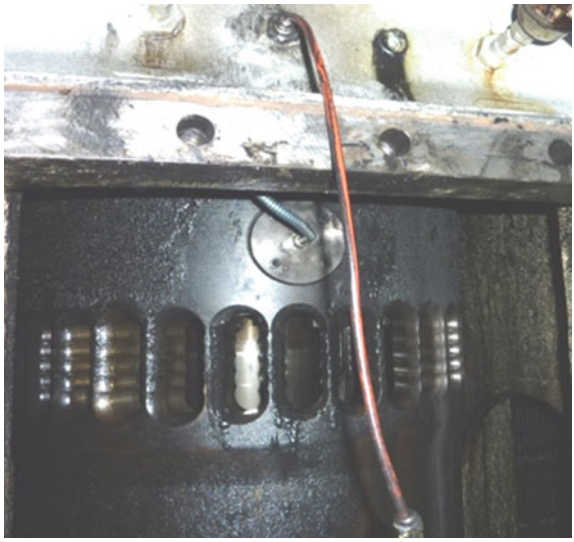


Fig. 7 Installation drawing of sensor

acquire the piston ring wear signal depend on the reciprocating of piston ring. Due to the real ship test conditions, the teat is done only under normal conditions of piston rings.

4.2 Comparison with the Experiment and the Results

Calculation condition setting with the same test conditions, the monitoring of the magnetic field strength is calculated by multiplying the obtained calibration of the sensitivity of each sensor, the simulation results can be converted to a voltage change. The simulation results were compared with the measured waveform. Sensor output voltage curve shown in Fig. 8a, the curves represent four piston moves past the sensor output voltage changes; Comparison Fig. 8a, b, the output signal waveforms found consistent calculation of the theoretical analysis and experimental testing the same law to verify the accuracy of the simulation model. Both the voltage waveform is not exactly the same numerical size that mainly because the actual test voltage amplification factors of the tune and calibration experiments vary the voltage magnification. And in the actual test engine piston rear wear rings, finite element modeling process in order to reduce the amount of computation, only established a piston head model.

Experiments of the real ship show that: modeling realistic, validated simulation models, methods and conclusions are correct. Using the model, it can be calculated that the relationship between the output voltage signals by theoretical models can be instead of the actual test. From the simulation we can get the relationship between different marine diesel piston ring wear and the sensor output voltage. Thereby a

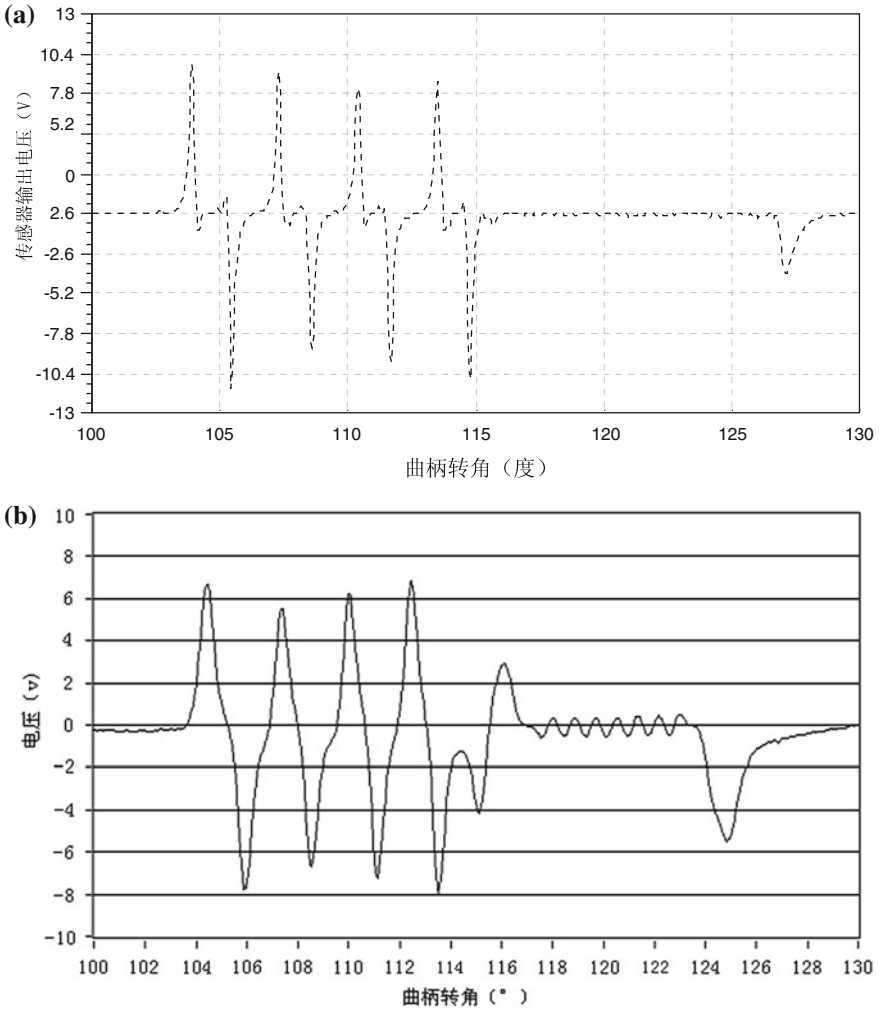


Fig. 8 Simulation result and test result of sensor

large number of piston ring wear tests can be reduced. It provided technical support for the realization of the state of marine diesel engine piston ring wear line monitoring project.

5 Conclusions

In this paper, for the study of marine diesel engine, the calculation of magnetic field three-dimensional finite element calculations, the development and calibration of the sensor, and the combination of the real ship research. Studies show that here is

single corresponding relationship between the amount of wear of piston rings and monitoring point field strength; The calibration tests of the piston ring wear monitoring sensor shows it the high sensitivity to meet technical requirements of the wide range of marine diesel engine piston ring wear and high resolution ratio; the finite element model, the correctness of calculation methods is verified, it shows that the calculation model can replace the different real ship tests. Binding studies of theoretical calculations and the actual hardware development provide technical support for realization on line monitoring piston ring wear of marine diesel engine.

References

1. Huang S (2005) The Analysis of Modern Marine Diesel Engine Failure. Dalian maritime university press, Dalian, pp 134–140
2. Yang J, Peng Z, Yu Y, Lu Y (2010) Research on monitoring method of piston ring wear based on magneto-resistive sensor. Trans CSICE 28(1):85–89
3. Peng Z (2010) Research on Monitoring Method of Piston Ring Wear Based on Magneto-Resistive Sensor for Marine Diesel Engine. Wuhan University of Technology, Wuhan
4. Wang Z, Hu B (2005) Discuss on the application of permalloy magneto resistance sensor. Acoust Electron Eng 79(3):47–49
5. Bai H (2006) Design of system for weak magnetic fields measurement based on thin-film magnetoresistance sensors. Instrum Tech Sens 10:30–31
6. Zhangming P, Jianguo Y, Qiaoying H (2011) Research on monitoring wear of piston ring based on magneto-resistive sensor for marine diesel engine. Measuring Technology and Mechatronics Automation, China
7. Zhao B, Zhang H (2009) Application of Ansoft 12 in the engineering electromagnetic. China Water Conservancy and Electricity Press, Beijing
8. Li Q (2002) Numerical calculation of electromagnetic fields and electromagnet design. TsingHua University Press, Beijing, p 17
9. Sifuentes E, Casas O, Pallas-Areny R (2007) Direct interface for magnetoresistive sensors. Instrumentation and Measurement Technology Conference–IMTC 2007 Warsaw, Poland, pp 1–6

Criteria and Performance Survey in Applying PAS 55 to Hong Kong Buildings and Plants

Samuel K.S. Fung and Peter W. Tse

Abstract Following the global trend of escalating customer expectation of services and products in the trade of Engineering Asset Management (EAM), a systematic and international engineering management system was developed within the Publicly Available Specification: Asset Management in Year 2004 (PAS 55:2004) by British Standard Institution (BSI). The International Standards Organization (ISO) has adopted its subsequent edition [1] for development of asset management series of international standards. This article highlights an application research served to conduct a tentative sampling survey on about 31 building and plant practitioners in EAM. On following their management systems mapped according to the PAS 55 framework, their O&M performances can be evaluated by an artificial intelligence based method, which is designed according to PAS 55 criteria, and used to establish models. The intelligent method makes use of the survey information to model the benchmarking levels of PAS 55 and the requirements for different categories of EAM practitioners. Section 4.4.6 Information Management of PAS 55:2008 was selected in this study. A questionnaire was designed to survey the performance of the local 31 building practitioners in EAM. The survey result has been adopted as a reference to virtual adoption levels at PAS 55. Practical means have also been revealed so that the building O&M practitioners can find their ways to accomplish a full recognition in EAM and vital references in benchmarking their performance with the world recognized EAM performance.

S.K.S. Fung (✉)
Karson Engineers Services Co. Limited, Kwai Chung, Hong Kong
e-mail: samuelfung@karson-eng.com
URL: http://www.karson-eng.com

P.W. Tse
The Smart Engineering Asset Management Laboratory (SEAM), Department of Systems Engineering and Engineering Management (SEEM), City University of Hong Kong, Kowloon Tong, Hong Kong
e-mail: Peter.W.Tse@cityu.edu.hk
URL: https://www6.cityu.edu.hk/seam

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_62

1 Introduction

The management framework of PAS 55 is compatible to interface with Plan-Do-Check-Act (PDCA) approach commonly adopted in ISO Management Systems. Whereas the technical elements of the 28 criteria as listed in Sections 4.1–4.7 of PAS 55:2008 may be elaborated in the use of various BSI, EN, IEC and ISO publications which are readily available for technical application in EAM trade. Variance of Engineering Asset Management requirements at various sites and countries on certain focus on local needs may develop even though they are different from the international standards developed under a common framework as PAS 55: 2008 which will be adopted as ISO 55000 in 2014. Resolutions of local professionals and international expertise for coordinated applications in meeting both the local and international management requirements may adopt mapping skills to be developed locally in any country wherever it demands.

EAM performance assessment substantially depends on relationship between terotechnology and various professional activities [2], that was made to definitions in BS3811 subsequently BS EN 60300:2011 as combination of engineering, management, financial and other trade practices applied to physical assets in pursuit of economic life-cycle cost. For those managers of non-EAM professionals, mathematical calculation of assessment scores on the EAM performance are suggested to be enhanced by Fuzzy Logic Averaging which applies to use of fuzzy sets and applications [3]. Linguistic variables of the EAM stakeholder assessments of non-EAM professionals on the EAM performance may thus be processed with linguistic modifiers with fuzzy sets and patterns of fuzzy memberships for computation of Fuzzy Logic Averaging, and this computation result will serve as the performance scores under the PAS 55 management framework. For simplicity to carry out modeling in this research in Hong Kong, EAM professional practitioners were invited to attend interviews and give answers directly to our questionnaire survey on their EAM cases instead of conducting a general survey which may need use of such Fuzzy Logic Averaging technique.

2 Method of Applied Research

An applied research is based on a Small-Medium-Enterprise (SME) consulting engineers services company (Karson Engineers Services Co Ltd) operating in the EAM sector in Hong Kong, and served to conduct a sampling survey on his company clients of 31 building and plant practitioners selected in the EAM. Selection of the sampling size was made in reference [4] that the size of 30 will be considered adequate for expected correlation greater than 0.5 which was the case in our supervised data mining. Details of 31 sampled EAM practitioners are listed in Table 1:

Due to time limitation of this application research, the tentative artificial intelligence mechanism used dedicated software of IBM SPSS (Statistics)® and MS

Table 1 The nature of the surveyed buildings and plants

ID no.	Details of reply maker (EAM practitioners)	Portfolio
1	Government related offices buildings (regional A&A)	Public services
2	Government related offices buildings (regional O&M works)	Public services
3	Government related housings (regional O&M works)	Public services
4	Government related housings (local estate and shopping arcade)	Public services
5	Government related housings (district estates and shopping arcades)	Public services
6	University campus buildings	Public services
7	University laboratories	Public services
8	International banking corporation 1—regional banking headquarters buildings	Commercial
9	International banking corporation 2—regional banking headquarters buildings	Commercial
10	International banking corporation 3—HK (retails and regional offices)	Commercial
11	International properties developer—commercial complex	Commercial
12	Local enterprise properties developer 1—commercial complex	Commercial
13	International banking corporation 1—Data centres and comms rooms	Commercial
14	International banking corporation 2—data centres and comms rooms	Commercial
15	Local enterprise properties developer 2—DATA CENTRES	Commercial
16	Telecom ISP corporation—data centres	Commercial
17	International banking corporation 3—data centres	Commercial
18	SME commercial building	Commercial
19	International banking corporation 1—regional staff residences	Residential
20	Residential court 1 (residential owners under Government subsidy scheme)	Residential
21	Residential court 2 (residential owners under Government subsidy scheme)	Residential
22	Residential development (private estate)	Residential
23	Residential building (residential owners from asset sale of international banking corporation)	Residential
24	Shopping arcades—real estates investment	Composite
25	Local properties group 1—retail shops	Composite
26	Local properties group 2—group offices and shops	Composite
27	Local caterers restaurants and fast food shops	Composite
28	Venue provider- centers and hotels	Composite
29	Industrial practitioner- local dockyards	Industrial
30	International industrial group-factory building	Industrial
31	SME industrial centre	Industrial

EXCEL® as demonstration on a selected part of Information Management (i.e. Section 4.4.6 Part a, b, c, and f) of PAS-55: 2008 for the modeling and moderation of levels of adoption of PAS-55. Surveying answers of Questionnaires Form containing 60 questions, i.e. Q1-Q60 were sought from the 31 IDs in between October 12 and March 13 as per the survey scheme as follows.

2.1 Focus Pattern of Questionnaires in This Application Research

Clause 4.4.6 Information Management of PAS 55:2008 for Engineering Asset Management (EAM) is selected for demonstration purpose of assessing 1 of 28 criteria which are adopted as Key Performance Indicators on complying PAS 55 as a base for future benchmarking development to the similar assessment for other 27 criteria. The Information Management is typically required as common in PAS 55, ISO 9001, ISO14001 and OHSAS 18001. This demonstration elaborated details according to the 4 of 6 Sub-Clauses in PAS 55:2008, i.e. 4.4.6(a)–(c) and 4.4.6(f) to conduct an application survey from 31 EAM practitioners on the following:

PAS-55 Clause 4.4.6(a) *Adequacy of Information Authorized for Use of Asset Management*

Q1–Q16 are survey on: Information Manager assignment; Information Structure establishment; Information essentials such as Asset Registers, Drawings, Contracts, Licenses, Legal Regulatory and Statutory Documents, Policies, Standards, Guidance Notes, Technical Instructions and Procedures, Operating Criteria, Performance and Condition Data, Tacit Knowledge, T&C and O&M Records; Control of information accuracy; Information to enable optimization and prioritization, assess financial benefits, determine operational and financial impact on unavailability or failure of the major operations, compare life cycle costs among alternatives, monitor details and expiry dates of licenses, warranties and certifications, etc., determine with costs of activities and replacements with track record of market prices, determine end of economic life of the major engineering asset with track records of paid rates, allow for performing financial analysis of planned income and expenditures, determine financial and resource impact on availability and performance over a contingency period if contingency plan is taken place, assess overall financial performance of the engineering assets, allow to perform risk analysis for operation and maintenance works, assure performance of statutory compliance with track records with respect to the rules.

PAS-55 Clause 4.4.6(b) Periodic Review and Revision to Maintain Adequacy of the Information Management System

Q17–Q34 are survey on: Consistent coded names of asset items information; Information to manage asset life cycles their legal and regulatory management requirements, describe assets, functions and systems being served, access planning and work O&M schedules; Information to give unique asset identification and asset registers, locations and spatial layout of assets, engineering data, design parameters, and drawings, vendor data for assets, testing and commissioning dates and data of assets, task risk assessments and control measures, task details of the last maintained /inspected and when they are next due, listing of overdue /outstanding tasks, historical record of planned and unplanned maintenance tasks performed, operational data including performance characteristics and design limits, financial data of available cost, cost of historical pm tasks, operating cost, downtime impact, replacement value, initial cost, etc., working programmes and schedules of works and settings (long and short terms), planning of asset possession, shutdown and outage, operating details of condition monitoring systems.

PAS-55 Clause 4.4.6(c) Allocation of Appropriate Roles and Responsibilities and Authorities in using the Information Management System

Q35–Q50 are survey on: Information accessible and available to all relevant personnel under monitoring and controls; Allocation of responsibilities and authorities for maintenance, access, archiving and disposal of information; Information maintenance, version control and assurance activities, generation, capture or importing of the identified items, ownership and maintenance demarcation where assets interface across a system or network of assets; Asset service requirements, conditions, and performance targets or standards; Requirements of key performance indicators; Criteria of non-conformance and the actions to be taken; Details of emergency plans, responsibilities and contacts; Information of asset build-up conditions and duty use, current tasks and planned works, materials, inventory, purchasing management systems, decision-supporting systems for optimization and life cycle costing models, service performance reporting systems, staff locations, scheduling and dispatch systems, capital expenditure planning and condition monitoring systems.

PAS-55 Clause 4.4.6(d) Assurance against Unintended Use of Obsolete Information in using the Information Management System, and Clause 4.4.6(e) Assurance of Archival Information retained for Legal or Knowledge Preservation in the Information Management System are dealt as political issues under a separate cover.

PAS-55 Clause 4.4.6(f) Assurance of Information Security with Back-up Recovery in the Information Management System

Q51–Q53 are survey on: Storing information items according to integrity, security and confidentiality; Performing management cycles of establishment, implementation, retention, and disposal of records; Monitoring effectiveness of record procedures, access controls and storage facilities and disposal.

The answering scores of Q1–Q53 are those the EAM Practitioner agrees or disagrees with the questionnaire statements by selecting scoring boxes of the following 5 scales, namely: 1 = Totally Adopted (91–100 %); 2 = Mostly Adopted (90–75 %); 3 = Generally Adopted (74–41 %); 4 = Slightly Adopted (40–11 %) and 5 = Not Adopted (0–10 %).

2.2 Conclusion Questionnaires for the 4 Clauses of 4.4.6 (a)–(c) and (f) of PAS 55:2008

(A) What Damages if Any if Part of the Information Management System of PAS-55 Not in Use for the O&M of the Engineering Asset

Q54–Q57 are survey on: Damages if any when PAS-55 Clause 4.4.6(a) Information Authorization for Use not adopted (Q54); if any when PAS-55 Clause 4.4.6(b) Periodic Review on Revision to Maintain Use not adopted (Q55); if any when PAS-55 Clause 4.4.6(c) Allocation of Roles and Responsibility on Use not adopted (Q56); if any when PAS-55 Clause 4.4.6 (f) Assurance and Back-up Recovery on Use not adopted (Q57).

More significant damages are regarded as higher importance to the ID EAM practitioners. The answering scores of Q54–Q57 are those the EAM Practitioner agrees or disagrees with the questionnaire statements by selecting scoring boxes of the following 5 scales, namely: 1 = Most significant; or 2, 3, 4, reducing towards 5 = Least significant.

(B) Overall Management Performance of the O&M of the Engineering Asset Management

Q58–Q60 are survey on: Evaluation of good practice in Cost engineering (Q58), Quality/Reliability Management (Q59), and Time /Efficiency Management (Q60).

The answering scores of Q58–Q60 are those the EAM Practitioner agrees or disagrees with the questionnaire statements by selecting scoring boxes of the following 5 scales, namely: 1 = Most satisfactory; or 2, 3, 4, reducing towards 5 = Least satisfactory.

2.3 Analysis of the Result of the Questionnaire Survey

For higher accuracy of the questionnaire answers collected, most of the sampled EAM Practitioners (ID Nos. 1–31) were interviewed by Q&A process upon availability of meetings within the period of 6 months ended in March 2013. All questionnaires were answered and analyzed as per the following sections.

3 Data Analysis of Application Research

The key data of questionnaire answers (Q1–Q60) of 31 sampled EAM practitioners (ID Nos. 1–31) were divided by 4 sets of SPSS analysis and input into IBM SPSS® for Standard Multiple Linear Regression as follows:

3.1 Use of SPSS for Standard Statistical Inference

For those managers of EAM professionals, survey of EAM performance ratings in linguistic variables are directly applicable in this Project under the PAS 55 framework. As variance of EAM performance ratings in auditing survey in form of questionnaire replies among EAM professionals are to be scientifically processed, engineering statistics, Multiple Linear Regression for the EAM performance evaluation and statistical modeling are adopted as typically described in text books of the following:

(a) Applied Linear Regression Models [5];
(b) Statistics Concepts and Controversies [6].

The performance assessment ratings may then be formulated as follows:

For Y = Score of satisfaction level of the overall sub-section of the PAS 55 criteria (dependent variable);

X_{Q1} = Score of adopting effective use of the sub-sectional requirement of the PAS 55 criteria (independent variable Q1);
↕
$X_{Q}n$ = Score of adopting effective use of sectional requirement of the PAS 55 criteria (independent variable Qn);

then

$$\mathbf{Y} = \mathbf{A} + \boldsymbol{\beta}_{\mathbf{1}*}\mathbf{X}_{\mathbf{Q1}} + \ldots + \boldsymbol{\beta}_{\mathbf{n}*}\mathbf{X}_{\mathbf{Q}}\mathbf{n} + \boldsymbol{\varepsilon} \tag{3.1}$$

where A is the regression constant; $\beta_1, \ldots, \beta_n$ are regression coefficients; and ε is residual of regression. When sampling of the regression reach significant level of confidence, ε would be considered as null and so Y will approximately become $\hat{Y}$ as the expected value of Y in the sampling of regression and modeling.

In questionnaires of surveys on auditing performance assessment for statistical inference and modeling, vast numbers of dependent variables and independent variables are too difficult for human calculations of matrices, and so statistical software computation is required in practice. The popularly available statistical software for questionnaire surveying is found as "Statistical Package for the Social Science (SPSS)" which can be readily sought for this project application research as IBM SPSS ® for MS Windows. Thus the statistical inferences and modeling analysis for this research of PAS 55 applications adopts the SPSS with the references [4] and [7].

SPSS Analysis 1 (Q54 vs. Q1–Q16 for performance of PAS 55 Section 4.4.6 (a)—Adequacy of Information Authorized)

The Performance Output (Dependent Variable) is Q54;

The Performance Predictors (Independent Variables) are Q1, Q2,…, Q16

The computation report on Explore and Regression (full report available on request) was discussed as follows:

Variables	Median values	Outliers IDs (respect to individual Qs) which need individual modeling	Significant correlations (overall Q54 to Q1–Q16) or (Q54 to individual Qs)
Q54	2	–	0.9 (Overall Q54 vs Q1-Q16)
Q1–Q16	3, 2, 2, 2, 3, 3, 3, 3, 2, 2, 3, 3, 2, 2, 3, 2.	2, 27 (Q1); 23 (Q4); 23, 29 (Q5); 9 (Q13); 23 (Q14).	0.5 (Q2); 0.4 (Q3); 0.5 (Q4); 0.4 (Q5); 0.3 (Q6); 0.5 (Q7); 0.4 (Q9); 0.5 (Q10); 0.3 (Q11); 0.4 (Q12); 0.6 (Q13); 0.7(Q15); 0.5 (Q16).

Significant coefficient(s) of the linear equation: $\beta_4 = 0.615$ for Q4
None of Independent Variables are mutually affecting on collinearity
Hypothesis tested as Associated with the Population at 95 % confidence level

SPSS Analysis 2 (Q55 vs. Q17–Q34 for performance of PAS 55 Section 4.4.6 (b)—Periodic Review and Revision of Information Adequacy)

The Performance Output (Dependent Variable) is Q55;

The Performance Predictors (Independent Variables) are Q17, Q18,.., Q34

The computation report on Explore and Regression (full report available on request) was discussed as follows:

Variables	Median values	Outliers IDs (respect to individual Qs) which need individual modeling	Significant correlations (overall Q55 to Q17–Q34) or (Q54 to individual Qs)
Q55	2,	–	0.9 (Overall Q55 vs Q17-Q34)
Q17–Q34	2, 2, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3,2, 2, 2.	29 (Q19); 7, 9 (Q25) 9, 23 (Q26); 7 (Q27); 7 (Q28); 6, 9, 23, 29 (Q31); 9 (Q32).	0.7 (Q17); 0.3 (Q20); 0.4 (Q21); 0.5 (Q22); 0.3 (Q24); 0.4 (Q26); 0.4 (Q29); 0.4 (Q30); 0.4 (Q34).

Significant coefficient(s) of the linear equation: A = 1.1, β_{17} = 0.4 for Q17, β_{19} = 0.6 for Q19, β_{25} = −0.8 for Q25, β_{31} = −0.3 for Q31
Independent Variables Q22, Q24, Q25, Q27 and Q28 are mutually affecting on collinearity
Hypothesis tested as highly Associated with the Population at 95 % confidence level

SPSS Analysis 3 (Q56 vs. Q35–Q50 for performance of PAS 55 Section. 4.4.6 (c)—Allocation of Roles and Responsibilities and Authorities in using Information)

The Performance Output (Dependent Variable) is Q56;

The Performance Predictors (Independent Variables) are Q35, Q36,.., Q50

The computation report on Explore and Regression (full report available on request) was discussed as follows:

Variables	Median values	Outliers IDs (respect to individual Qs) which need individual modeling	Significant correlations (overall Q56 to Q35–Q50) or (Q56 to individual Qs)
Q56	2	–	0.9 (Overall Q56 vs Q35-Q50)
Q35–Q50	2, 2, 3, 3, 2, 3, 3, 2, 2, 2, 2, 2, 3, 2, 2, 2.	5, 12, 13, 23, 28, 29 (Q35); 18, 31 (Q37); 7, 29 (Q50).	0.5 (Q35); 0.5 (Q36); 0.5 (Q37); 0.5 (Q38); 0.4 (Q39); 0.5 (Q40); 0.6 (Q41); 0.6 (Q42); 0.3 (Q43); 0.5 (Q44); 0.6 (Q45); 0.3 (Q46); 0.6 (Q47); 0.5 (Q48); 0.7 (Q50).

Significant coefficient(s) of the linear equation: A = 1.0, β_{50} = 0.6 for Q50
Independent Variables Q38 and Q39 are mutually affecting on collinearity
Hypothesis tested as highly Associated with the Population at 95 % confidence level

SPSS Analysis 4 (Q57 vs. Q51-Q53 for performance of PAS 55 Section. 4.4.6 (f)—Assurance of Information Security and Back-up Recovery of Information)

The Performance Output (Dependent Variable) is Q57;

The Performance Predictors (Independent Variables) are Q51, Q52, Q53

The computation report on Explore and Regression (full report available on request) was discussed as follows:

Variables	Median values	Outliers IDs (respect to individual Qs) which need individual modeling	Significant correlations (overall Q57 to Q51–Q53) or (Q57 to individual Qs)
Q57	2	18, 19, 21, 23	0.4 (Overall Q57 vs Q51-Q53)
Q51–Q53	2, 2, 2.	23 (Q52);	0.3 (Q51); 0.4 (Q52); 0.4 (Q53).

Significant coefficient(s) of the linear equation: A = 1.1
None of Independent Variables are mutually affecting on collinearity
Hypothesis tested as Not Associated (non-consistence) with the population at 95 % confidence level

SPSS Analysis 5 (Q58 vs. Q54-Q57 for Cost satisfaction vs Information Management performance of PAS 55 Section. 4.4.6(a)–(c) and (f))

The Performance Output (Dependent Variable) is Q58;

The Performance Predictors (Independent Variables) are Q54, Q55, Q56 and Q57

The computation report on Explore and Regression (full report available on request) was discussed as follows:

Variables	Median values	Outliers IDs (respect to individual Qs) which need individual modeling	Significant Correlations (overall Q58 to Q54 – Q57) or (Q58 to individual Qs)
Q58	2	–	0.4 (Overall Q58 versus Q54–Q57);
Q54–Q57	2, 2, 2, 2.	18, 19, 21, 23 (Q57).	(null on Q54), (negative 0.7 at significant level of 0.07 on Q55); (negative 0.7 at significant level of 0.07 on Q56); (null on Q57).

Significant coefficient(s) of the linear equation: A = 1.3
None of Independent Variables are mutually affecting on collinearity
Hypothesis tested as Associated with the population at 95 % confidence level

SPSS Analysis 6 (Q59 vs. Q54–Q57 for Quality satisfaction vs Information Management performance of PAS 55 Section. 4.4.6(a)–(c) and (f))

The Performance Output (Dependent Variable) is Q59;

The Performance Predictors (Independent Variables) are Q54, Q55, Q56 and Q57

The computation report on Explore and Regression (full report available on request) was discussed as follows:

Variables	Median values	Outliers IDs (respect to individual Qs) which need individual modeling	Significant Correlations (overall Q59 to Q54 – Q57) or (Q59 to individual Qs)
Q59	2	–	0.6 (Overall Q59 versus Q54–Q57)
Q54–Q57	2, 2, 2, 2,	18, 19, 21, 23 (Q57)	0.4 (Q54); 0.5 (Q55); 0.5 (Q56); 0.4 (Q57).

Significant coefficient(s) of the linear equation: none
None of Independent Variables are mutually affecting on collinearity
Hypothesis tested as Associated with the population at 95 % confidence level

SPSS Analysis 7 (Q60 vs. Q54–Q57 for Time satisfaction vs Information Management performance of PAS 55 Section. 4.4.6(a)–(c) and (f))

The Performance Output (Dependent Variable) is Q60;

The Performance Predictors (Independent Variables) are Q54, Q55, Q56 and Q57

The computation report on Explore and Regression (full report available on request) was discussed as follows:

Variables	Median values	Outliers IDs which need exclusion in need	Significant correlations (overall Q60 to Q54–Q57) or (Q60 to individual Qs)
Q60	2	–	0.4 (Overall Q60 vs Q54-Q57)
Q54–Q57	2, 2, 2, 2.	18, 19, 21, 23 (Q57).	0.4 (Q54); 0.4 (Q56).

Significant coefficient(s) of the linear equation: A = 1.4

None of Independent Variables are mutually affecting on collinearity

Hypothesis tested as Not Associated (non-consistence) with the population at 95 % confidence level

3.2 The Use of Least Square Euclidean Distance for Data Mining and Moderation

To find moderated operating points in supervised clusters of assessed performance levels of surveyed models, target performance of building plants within the same cluster will be moderated by Artificial Intelligence. In practice of application research, finding the moderated operating points by computing averaged Euclidean Distance of operating points of several similar buildings within supervised clusters are introduced. On taking the reference [8] that use of supervised data may perform linear discrimination of data mining as basic operation of support vector machine on k-nearest neighbor algorithm. On use of k-means clustering by taking the initial averaged point as the first centroid for the first round least square Euclidean Distance calculation. Then take the usually admitted value of K = 3 for seeking the least square distance of the nearest centroid which is the recommended optimized operating point among the existing surveyed points. For a trial operation of this algorithm, Use of ID Nos. 1–7 (Public Sector of EAM Practitioners) for seeking the optimum point with respect to Q54 versus Q1–Q16 is illustrated in Figs. 1, 2 and 3 on using K-distance computation procedures shown below:

(a) List all scores of Q1–Q16 for Q54 with respect to ID 1–ID 7 (Public Services EAM Practitioners, selected as a Supervised Learning Cluster for this demonstration);

(b) Compute K-Distance (Sum of Squares of Q1–Q16 scores) for individual IDs that is for overseeing relative correlations among IDs;

ID No.	Scores 1-5 with respect to Questionnaires Q1-16																K-distance (Sum of Squares of Q1-16)	Check Q
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16		Q54
1	1	1	2	1	2	3	3	3	1	2	3	3	3	2	1	1	76.0	1
2	2	2	2	2	3	3	3	3	2	3	3	3	3	3	3	3	119.0	3
3	1	1	1	1	1	2	3	3	3	3	4	2	3	3	3	2	96.0	2
4	5	5	3	2	3	3	3	3	2	3	3	4	3	3	3	2	168.0	2
5	4	3	3	3	4	4	4	4	1	2	3	3	3	3	4	2	168.0	3
6	3	2	2	3	2	2	3	2	1	1	1	3	2	4	2	1	84.0	2
7	2	4	4	3	3	2	2	3	3	2	4	2	4	3	4	3	154.0	4
Centroid 1 (average)	2.57	2.57	2.4	2.14	2.57	2.7	3	3	1.86	2.29	3	2.86	3	3	2.86	2	111.7	2.4285714
Least Square Dist.																		
LSD C1-ID1	2.47	2.47	0.2	1.31	0.33	0.1	0	0	0.73	0.08	0	0.02	0	1	3.45	1	28.5	2.0408163
LSD C1-ID2	0.33	0.33	0.2	0.02	0.18	0.1	0	0	0.02	0.51	0	0.02	0	0	0.02	1	1.5	0.3265306
LSD C1-ID3	2.47	2.47	2	1.31	2.47	0.5	0	0	1.31	0.51	1	0.73	0	0	0.02	0	27.9	0.1836735
LSD C1-ID4	5.9	5.9	0.3	0.02	0.18	0.1	0	0	0.02	0.51	0	1.31	0	0	0.02	0	71.7	0.1836735
LSD C1-ID5	2.04	0.18	0.3	0.73	2.04	1.7	1	1	0.73	0.08	0	0.02	0	0	1.31	0	16.0	0.3265306
LSD C1-ID6	0.18	0.33	0.2	0.73	0.33	0.5	0	1	0.73	1.65	4	0.02	1	1	0.73	1	24.9	0.1836735
LSD C1-ID7	0.33	2.04	2.5	0.73	0.18	0.5	1	0	1.31	0.08	1	0.73	1	0	1.31	1	19.2	2.4693878
Centroid 2 (at K=3 of LSD)	2.67	3	3	2.67	3.33	3	3	3.33	2	2.33	3.33	2.67	3.3	3	3.67	2.67	140.8	3.3333333
Least Square Dist.																		
LSD C2-ID1	2.78	4	1	2.78	1.78	0	0	0.11	1	0.11	0.11	0.11	0.1	1	7.11	2.78	95.9	5.4444444
LSD C2-ID2	0.44	1	1	0.44	0.11	0	0	0.11	0	0.44	0.11	0.11	0.1	0	0.44	0.11	2.9	0.1111111
LSD C2-ID3	2.78	4	4	2.78	5.44	1	0	0.11	1	0.44	0.44	0.44	0.1	0	0.44	0.44	80.1	1.7777778
LSD C2-ID4	5.44	4	0	0.44	0.11	0	0	0.11	0	0.44	0.11	1.78	0.1	0	0.44	0.44	49.6	1.7777778
LSD C2-ID5	1.78	0	0	0.11	0.44	1	1	0.44	1	0.11	0.11	0.11	0.1	0	0.11	0.44	6.8	0.1111111
LSD C2-ID6	0.11	1	1	0.11	1.78	1	0	1.78	1	1.78	5.44	0.11	1.8	1	2.78	2.78	62.8	1.7777778
LSD C2-ID7	0.44	1	1	0.11	0.11	1	1	0.11	1	0.11	0.44	0.44	0.4	0	0.11	0.11	5.9	0.4444444
Thus the ID No.2 is the selected model of the least square support vector(K=3) in the public sector section.																		
The aimed scores of Q1-Q16 and Q54 may refer to Centroid 2.																		

Fig. 1 Least square euclidean distance calculation (K = 3) for moderation

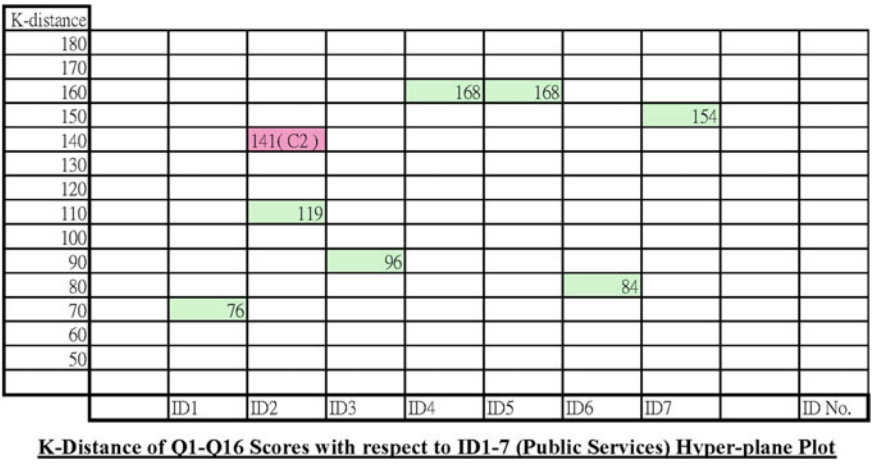


Fig. 2 Plot of moderated operating point (centroid no. C2) on K-distance of ID 1–7 and C2

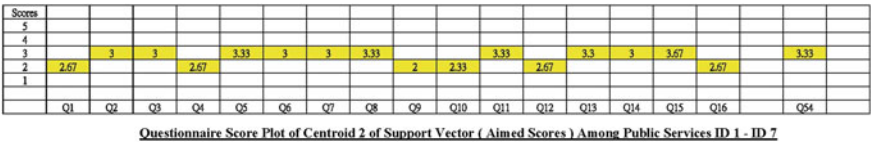


Fig. 3 Plot of moderated operating point (centroid no. C2) on aimed score of Q1–Q16 and Q54

(c) Attempt an initial Centroid ID (labeled as C1) by taking averaged scores of Q1–Q16 and Q54 among ID1–ID7;
(d) Compute relative K-Distance (Sum of Square difference) of Q1–Q16 and Q54 between Centroid ID–C1 and individual IDs (ID1–ID7);
(e) For K = 3, look for the first three minimum relative K-Distance (in this demonstration, ID 2, 5 and 7, for further computation (i.e. Go back to (c) until the minimum relative K-Distance remains for the same ID);
(f) Attempt the 2nd Centroid ID (labeled as C2) by taking averaged scores of Q1-Q16 and Q54 among ID2, ID5 and ID7;
(g) Compute relative K-Distance (Sum of Square difference) of Q1–Q16 and Q54 between Centroid ID–C2 and individual IDs of K = 3 (ID2, ID5 and ID7);
(h) As it was found that ID 2 has the first minimum relative K-Distance to Centroid C1 and remains the same position with Centroid C2. This least square distance constitutes that ID2 should be the selected model among Public Services EAM practitioners and Centroid ID C2 may be regarded as the targeted Public Services EAM practitioners in trade.

4 Conclusion and Further Research Opportunities

The outcome of this research is demonstrated by use of PAS-55 Plan-Do-Check-Act management framework that part of O&M performance of EAM practitioners in trade are typically able to be evaluated according to the PAS-55 criteria as Key Performance Indicators with mapped weighting survey factors for benchmarking to be developed in categorized use. An artificial intelligence mechanism such as statistical software and vector data mining computation can be used to determine tentative models and moderated benchmarking levels based on adoption of common PAS-55 criteria for guidance applications. Whereas further computation may adopt MatLab and LS-SVM Lab [9] to operate the Fuzzy Logic Averaging and Support Vector Machine in future research for fully automated artificial intelligence management systems with accumulating EAM data and different weighting factors for modeling and optimization of O&M operations and maintenance for desired outcome of different categories of EAM practitioners in trade.

References

1. PAS 55:2008, British Standards Institution (2008) Publicly available specification -asset management
2. Armstrong J (1993) Maintaining building services- a guide for managers. Mitchell, London, pp 24–25
3. Bejadziev G, Bojadziev M (1999) Fuzzy logic for business, finance, and management. World Scientific Publishing, Singapore
4. Dewberry C (2004) Statistical methods for organizational research- theory and practice. Psychology Press, London, pp 234–253
5. Kutner MH, Nachtsheim CJ, Neter J, Li W (2005) Applied linear regression models. McGraw-Hill, New York
6. Moore D, Notz W (2006) Statistics concepts and controversies. W.H. Freeman & Company, New York
7. Shannon D, Davenport M (2001) Using SPSS® to solve statistical problems. Merrill Prentice HallUpper Saddle River
8. Kaori S, Kazuaki T (2009) Illustration, Probabilities and statistics. Nitto Shoin Honsha Co. Ltd. c/o Tuttle-Mori Agency Inc., Tokyo
9. De Brabanter K, *P.* Karsmakers P, F. Ojeda F, Alzate C, De Brabanter J, Pelckmans K, De Moor B, Vandewalle J, Suykens JAK (2011) LS-SVM lab toolbox. Katholieke Universiteit Leuven, Belgium. http://www.esat.kuleuven.be/sista/lssvmlab

Competency Enhancement Model of Physical Infrastructure and Asset Management in Compliance with PAS-55 for Hong Kong Automotive Manufacturing Engineers

K.K. Lee, Raymond M.Y. Shan, Horace C.H. Leung and Joseph W.H. Li

Abstract To cope with the pitfall induced by the quick growth rate in asset management of automotive components manufacturing, an industry-wide professional competence development programme was initiated by SAE-HK and implemented by HKPC to upgrade the production and engineering asset management capability of Hong Kong automotive manufacturing engineers in order to boost the overall operation quality and efficiency of the industry. Through the comprehensive programme including the formulation of a human resources competence model for the industry, identification of major facility engineering and optimization tools in PAS-55 including Condition Assessment Technique (CAT); Machine Capability Index (MCI), Maintenance Analysis and Management (MAM) and Facility Management Information System (FIMS); PAS-55 system trial run and tools application at pilot companies; and the compilation of a best-in-class training and PAS-55 system implementation manual, local automotive component engineers are practically equipped with appropriate tools to understand the risks their businesses face, and the factors associated with facility optimization and prioritization.

Keywords PAS 55 · Human resources development · Hong Kong automotive manufacturing engineers

K.K. Lee (✉)
HKPC Building, 78 Tat Chee Avenue, Kowloon, Hong Kong
e-mail: kklee@hkpc.org

R.M.Y. Shan · H.C.H. Leung · J.W.H. Li
Materials and Manufacturing Technology Division, Hong Kong Productivity Council, Kowloon, Hong Kong

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_63

1 Introduction

1.1 The Chinese and Hong Kong Automotive Manufacturing Industries

Through years of double-digit expansion, Mainland China surpassed Japan to become the largest car producer in 2009 and the current Chinese market size is almost twice the size of the USA or Japan, far larger than any European country. Yet, the growth potential is still enormous as still less than 5 people in 1000 own an automobile. According to the China Association of Automobile Manufacturers, the overall automobile sales in Mainland China is expected to increase by 7 % to 20.6 million in 2013, safely securing her rank No. 1 in the world [1].

While the Chinese automotive industry and market grows by leaps and bounds, Hong Kong automotive components manufacturers have grasped a golden chance of enhanced special access to the high potential market in the light of CEPA and WTO commitments. The strong growth of the Chinese market has been leading the growth of the Hong Kong automotive industry, giving to the rise of a number of Hong Kong automotive manufacturing enterprises through the advancement in technological competence and the business scale.

1.2 More Than Fundamentals—Physical Infrastructure and Asset Management

According to Mr. Gordan Chan [2], ex-president of the Hong Kong Auto Parts Industry Association, in 2007 the Hong Kong automotive manufacturing industry consisted of 400-odd enterprises, of which about 30 were considered as "tier-one" which directly deal with high-end car manufacturers, while the remaining "tier-two" and "tier-three" operated in an OEM capacity. However, no matter the size and position of automotive manufacturers along the automotive supply chain, all suppliers have to get over dozens of stringent technical requirements on safety, reliability and product quality set by the automotive manufacturers and the upper tier customers to gain the entry ticket within the automotive manufacturing industry.

Through traditional engineering training, Hong Kong engineers are technically fit in automotive engineering, manufacturing engineering and facility design. However, inadequate training on equipment and machinery maintenance in traditional engineering disciplines could lead to incredibly increase in tools replenishment and equipment maintenance cost, extraordinary quick deterioration of working environment and potential pitfall in non-compliance with the requirement on Infrastructure Management and Work Environment Management stated in ISO/TS 16949, also the expectation of automotive manufacturers and upper tier customers.

1.3 PAS 55 and Asset Management Under the Spotlight

PAS 55 has been widely recognised as a significant step on the road of asset management. Many organisations and companies worldwide showed their interest in developing the specification and are actively wide spreading its use within the organizations. With the help and guidance of PAS 55, the Hong Kong based China Light and Power reported a 90 % reduction in system losses, while meeting a 20 % growth in demand and reducing customer charging tariffs by 40 %. Due to the ever success of PAS 55, the International Standards Organisation (ISO) has accepted PAS 55 as the basis for the development of the series of international standard ISO 55,000, which turns the best practice on asset management internationally for global excellence enhancement.

PAS 55 is a general standard for managing physical assets which is particularly relevant. It is deliberately structured to follow the design of other international standards including ISO 9000 and the Deming Plan-Do-Check-Act cycle of continual improvement. It also introduces the need for a number of essential tools to ensure alignment, integration and sustainability of efficient and effective asset management activities.

In 2008, PAS 55 was updated with input from 50 organizations in 10 countries, representing 15 industry sectors. It is increasingly recognized as a generically applicable definition of good practices in the whole life cycle, optimized management of physical assets. Comprising two parts, Part 1—Specification for the optimized management of physical infrastructure assets and Part 2—Guidelines for the application of PAS 55-1, it offers a 28-point checklist of requirements for an effective asset management system, defined terms and practical guidance on the implementation of the standard [3, 4].

2 Industry-Wide Human Resources Analysis and Enhancement

2.1 Urgent Needs of the Hong Kong Automotive Manufacturing Industry

ISO/TS 16949 has not been new to the Hong Kong automotive components manufacturers since 2002. The Hong Kong automotive manufacturing sector grows very quickly due to the growth of the Chinese market with the tangible support from the HKSAR Government. To keep the ball rolling, automotive components manufacturers must have effective management processes to maintain the high quality and reliable automotive components at a competitive cost, which is primarily dependent on the effective function of their manufacturing facilities and the stewardship of the physical assets such as production equipment, manufacturing

plant, auxiliary peripheral equipment, testing facilities, logistic facility, software programmes and system back up, etc.

According to the requirement of ISO/TS 16949, manufacturers shall determine, provide, and maintain the infrastructure needed to achieve product conformity and require a formal approach for infrastructure and facility management including contingency plan. Therefore, formal approach is explicitly required to infrastructure and facility management for automotive component manufactures in order to meet industry expectation. With a formal facility management system in place, automotive components manufacturers would be able to better understand the risk that their businesses face and factors associated for facility optimisation and prioritisation in order to achieve mutual benefits.

2.2 Human Resources Analysis and Modeling

To satisfy the industry's needs and to cope with the abovementioned problems, SAE-HK proposed an industry-wide professional competence development programme with Hong Kong Productivity Council (HKPC) to seek governmental funding support through the Professional Services Development Assistance Scheme (PSDAS) from the Commerce and Economic Development Bureau (CEDB). The proposed professional competence development programme targeted mainly on automotive manufacturing engineers; starting with the desktop search and analysis of the capability of human resources on asset management, and following by sample forms and tools design, train-the-trainers programme, practical on-site implementation trials, industry-wide enhancement training and training manuals compilation.

The human resources capability desktop search and analysis indicates the direction and the framework of the entire project. The aim of this phase is to identify the skillset of engineers within the automotive manufacturing sector based on the traditional engineering training at universities and technical colleges. Based on the results, the main targets of the professional competence development programme are identified. The results of the study of skillsets possessed by engineers from different engineering disciplines within this sector are shown in Fig. 1.

The automotive manufacturing sector is very closely linked with manufacturing engineering and mechanical engineering which focus on automotive parts and components design and manufacturing processes. These were the major revenue-generating areas where Hong Kong manufacturers focused on. Computer engineering and information engineering serve mainly as a business supporting role in the field of information technology within the sector, while civil engineering and electrical engineering and for the construction of manufacturing infrastructure and plant construction. Electronic engineering is the least relevant as the production of automotive signaling systems or other electronic automotive control systems are not typical to Hong Kong automotive manufacturers.

Fig. 1 Technical knowledge and skillset profile analysis for engineers from different engineering disciplines within the automotive manufacturing sector

Therefore, the main participants of our programme were automotive manufacturing engineers from the manufacturing and mechanical engineering background who had extensive exposure to expensive production equipment and necessary physical infrastructure. The programme structure was then formulated in accordance to their skillset they built up from traditional engineering training and the gap between the world class asset management best practice.

2.3 Comprehensive Series of Professional Competence Development Programmes

In the traditional manufacturing engineering and mechanical engineering training, the provision of courses on asset management and other similar subjects is comparatively generic and theoretical. There is no practical training on asset management for the automotive manufacturing sector offered by any local engineering institutions. Seeing this, 13 topics were chosen based on the human resources capability analysis to provide a comprehensive understanding on PAS 55 for the

Table 1 The training series within the professional competence development programme

Module	Topic
M1	Understanding of PAS-55 optimized facility management
M2	Facility engineering and optimization tool on life-cycle cost analysis (LCCA)
M3	Facility engineering and optimization tool on demand forecasting and management (DF&M)
M4	Facility engineering and optimization tool on machine capability index (MCI)
M5	Facility engineering and optimization tool on condition assessment and performance monitoring (CAPM)
M6	Facility engineering and optimization tool on risk assessment and management (RAM)
M7	Facility engineering and optimization tool on optimised decision-making (ODM)
M8	Facility engineering and optimization tool on maintenance analysis and management (MAM)
M9	Facility engineering and optimization tool on facility management information system (FMIS)
M10	Facility engineering and optimization tool on internal audit of risk-based management system of PAS55
M11	Facility engineering and optimization tool on facility management information system (FMIS)
M12	Facility engineering and optimization tool on documentation Development for facility management system
M13	Facility engineering and optimization tool on continual improvement

automotive manufacturing engineers. The details of the training series are shown in Table 1.

The first part of the training series was to train the trainers. 25 local engineers were identified and selected as the participants in the train-the-trainer section to learn the theories behind the asset management tools and to obtain practical knowledge on the application of asset management tools. The participants were then required to trial apply the facility engineering and optimization core tools on-site in their own factories so as to obtain hands-on experience. Then the 25 local engineers held the training sessions, with all 13 modules inclusive, in Mainland China to transfer their knowledge and share their experience to 75 Hong Kong engineers stationed in Mainland China.

Sample forms and tools were provided to facilitate the lectures and a practical session was arranged so that they could use the tools and knowledge in actual situations. All the course contents and the sample tools were compiled in the implementation manual to enhance the efficiency and effectiveness of knowledge transfer. One of the sample designed is shown in Fig. 2 and the cover of the implementation manual is shown in Fig. 3.

設備潛在失效模式及效應分析 (MFMEA)

__ 系統 _____ MFMEA 編號 _____

__ 次系統 _____ 頁數 第___頁，共_____頁

__ 零組件 _____ 設計責任 _____ 編制人 姓名_____ 部門_____ 職員編號_____

產品型號/計劃代號 _____ 關鍵日期 _____ DFMEA日期 (初稿)_____ (修訂)_____

跨部門小組 _____

項目 / 功能	潛在失效模式	潛在失效之效應	嚴重性	等級	潛在原因/失效機制	發生頻率	現行設計控制		難檢度	風險優先數	建議行動	責任與目標完成日期	行動結果			
							預防性	採測性					已採取行動	嚴重性	發生度	難檢度	風險優先數

Fig. 2 Sample form for MFMEA [5]

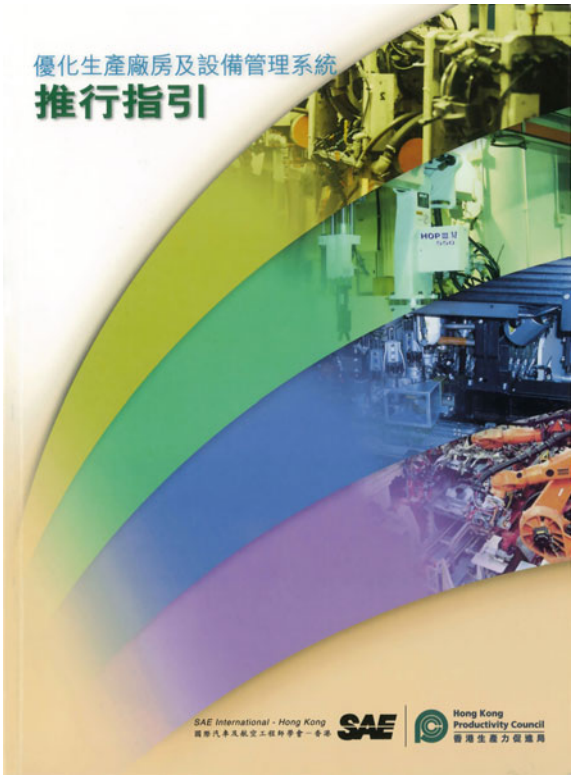


Fig. 3 The cover of the implementation manual [6]

3 Discussion

3.1 Effectiveness of the Professional Competence Development Programme

After the 28.5-day training series, the local automotive manufacturing engineers participated in the programme had shown their competence on PAS 55 through the successful application of various technical tools on asset management and the holding of teaching sessions in the PRD region. Apart from the direct beneficiaries at a size of 25 locally trained trainers and 75 Hong Kong engineers stationed in Mainland China, over 1300 engineers were benefitted by the implementation manual which contained the implementation guide, the set of sample facility engineering and optimisation tools as well as the set of PAS 55 process-based procedures.

3.2 Seed Driving Force of the Application of the Best Practice

Before the local trainers held their sessions in the PRD region, they had to trial implement PAS 55 according to the tools and skills taught during the lectures held by overseas speakers. The application skills were well proved to be practical and applicable through the two training sessions, which was commented as a very good arrangement for the local trainers to practice the newly learnt skills in the industry. From the attained performance of the local automotive manufacturing engineers, they could surely become the driving source to fasten the advancement of the industry through local knowledge sharing.

4 Conclusion

Within the automotive manufacturing industry, it has been conjoining, banding to pool resources since 1996 [2] The extensive growth in size and technology level requires high level on asset management, which is exactly addressed by PAS 55. In the foreseeable future, the development trend within this industry will drive a ballooning need for talents with PAS 55 or relevant asset management knowledge. The professional competence development programme showed successful knowledge transfer not only from the overseas speakers to the Hong Kong automotive manufacturing engineers, but also from these trainers to those engineers stationed in Mainland China. Through the provision of sample forms and tools specified in PAS 55, the trained engineers could pick up the skills on asset management quickly and able to apply those skills in actual situation. It is hoped that programmes in similar structures could be held in other industries so that the professional asset

management knowhow could widely be spread amidst Hong Kong industries, well supporting the growth of the industries grasping the golden chance provided by the Mainland market.

References

1. Russell F (2013) China's auto sales to rise by 7 % to record in '13, industry group says. Forbes: staff. http://www.forbes.com/sites/…/chinas-auto-sales-to-rise-by-7-to-record-in-13-industry-group-says/
2. The Federation of Hong Kong Industries (2007) Hongkong Industrialist. Hong Kong, China, pp 14–23
3. The Institute of Asset Management (2008). PAS 55-1:2008 asset management part 1: specification for the optimized management of physical assets. British Standards Institution, UK
4. The Institute of Asset Management (2008) PAS 55-1: 2008 asset management part 2: guidelines for the application of PAS 55-1. The British Standards Institution, UK
5. Hong Kong Productivity Council (2009). 優化生產廠秀及設備管理系統推行指引. Hong Kong, China, pp 2–30

Evaluation of Engineering Asset Acquisitions in EAM Based on DEA

Wei Liu, Wen-bing Chang and Sheng-han Zhou

Abstract This paper focuses on the process of acquisition of the engineering asset life cycle management and uses DEA (Data Envelopment Analysis) Model to evaluate the alternatives (decision making units, DMU). Firstly, based on the principles of the fuzzy clustering and rough set, we give an introduction to the object-weight-constrained DEA model. Then, Given the life cycle cost and taking the civil aircraft as an example, we take the cost of the acquisition, the cost of operation and maintenance, the cost of retirement disposal as the input variables and main performance parameters of the civil aircraft which include Wind Loading and Thrust-to-weight Ratio as the output variables, then, with the DEA linear programming and the use of LINGO, we can easily get the efficient DMUs and provide optimization suggestions for the inefficient DMUs. Also, we can get the scale benefit and technology availability of the DMUs respectively. Finally, we analyse the results and get the conclusion that the results meet the fact in the airlines.

1 Introduction

Engineering asset management which is to manage the tangible assets through its lifetime [1] including the Acquisition, Operation and Maintenance and the Retirement can make the assets more supportable for the realization of the organization's delivery strategy. Asset acquisition undoubtedly plays a vital role in it which will lay a solid foundation in the whole process. Many models can be used for the evaluation of the acquisition alternatives such as the Analytic Hierarchy Process (AHP), the Grey Model, the Fuzzy Model etc., but some of them are dependent on the subjective judgement and can only be used for the evaluation.

W. Liu (✉) · W.-b. Chang · S.-h. Zhou
School of Reliability and System Engineering, Beihang University, Beijing, China
e-mail: jancomeon@126.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_64

With the Object-Weights-Constrained DEA model, we can not only get an object view of the weight sequence but also evaluate the acquisition alternatives from quantitative aspect, also, some optimizations may be made for the alternatives.

2 DEA Model

With the principles of the fuzzy clustering and rough set, we get the object weight constraint of the output/input variables, and then, we integrate the constraints with the C^2R Model and get the Object-Weights-Constrained DEA model.

2.1 Weight Constraints Setting

With the fuzzy clustering and the importance principles of the rough set, we can get make full use of the original data and get the sequence result from that without any subject factors [2].

2.1.1 Data Calibration

In this step, we mainly get the fuzzy equivalent matrix from the original data.

Suppose that the set of influential factors is $(x_1, x_2, \ldots x_n)$, and we can get the original data matrix.

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} \tag{1}$$

After data standardization (0–1), we can get the below data matrix for analysis.

$$Y^* = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{pmatrix} \tag{2}$$

Then, with Max/Min Method (see 3), we can get the fuzzy similar matrix $R = (r_{ij})_{n\times n}$.

$$r_{ij} = \frac{\sum_{k=1}^{n} (y_{ik} \wedge y_{jk})}{\sum_{k=1}^{n} (y_{ik} \vee y_{jk})} \tag{3}$$

With the self-squared method, we can change the fuzzy similar matrix into the fuzzy equivalent matrix $R^* = (r_{ij})_{\mathrm{n}\times n}$ which shows the fuzzy equivalent relationships among the entities and meets the characteristics conditions of self-reverse, symmetry and transfer.

2.1.2 Fuzzy Clustering

Firstly, with the different fuzzy confidential level α set, we can set t_{ij} to 1 (If $r_{ij} \geq \alpha$) or 0 (if $r_{ij} < \alpha$) and get the new matrix $T_{\mathrm{k}} = (t_{ij})_{n\times n}$ for each specific α. By distributing the same value into the same cluster, we get the result set $C_i, i = 1, 2, \ldots k$ including the cluster group and member of each group for different values of α.

Then, by deleting a single factor from the all influential factors and repeating the previous step, we can get the result set $C_i^{'}, i = 1, 2, \ldots k^{'}$.

2.1.3 Grading

The resulting cluster only gives a group of entities with similar or the same characteristics, referring to the principles of the rough set, we can decide the importance of a specific factor $r_{i\alpha}$ by following the below principles.

$$r_{i\alpha} = \frac{|POS_{C_i}(C_i^{'})| - C_i^{'} \cap C_i}{|U|} + \frac{|POS_{C_i^{'}}(C_i)| - C_i^{'} \cap C_i}{|U|} \tag{4}$$

In which, $|POS_{C_i}(C_i^{'})|$ is the number of positive region of $C_i^{'}$ referring to C_i for a specific α and U is the total number of the sample. And then, we can get the importance of an factor with the below equation.

$$\bar{r} = \frac{1}{n} \sum r_{i\alpha} \tag{1.5}$$

In which, n represents the number of the classification for the α. Finally, we get the importance sequence of all the influential factors.

Note When there're many DMUs whose efficiency are 1, $\bar{r}$ can also be used to decide the best DMU by calculating the Euclidean distance between the real weights of the DMU factors and the $\bar{r}$.

2.2 Modelling

In this paper, we build the object-weights-constrained C^2R Model and BC^2 Model to analyze the productive efficiency and scale efficiency of the civil aircraft respectively. It can be represented as follows [3]:

$$(P^I_{C^2R})^{\varepsilon}\begin{cases}\max \theta_1 = \mu^T y_0 \\ \omega^T x_j - \mu^T y_j \geq 0, j = 1, \ldots n \\ \omega^T x_0 = 1 \\ \omega_i \leq \omega_j, 1 \leq i, j \leq n \\ \mu_i \leq \mu_j, 1 \leq i, j \leq n \\ \omega \geq \varepsilon e, \mu \geq \varepsilon \hat{e}\end{cases} \tag{6}$$

$$(P^I_{BC^2})\begin{cases}\max \theta_2 = (\mu^T y_0 - \mu_0) \\ \omega^T x_j - \mu^T y_j + \mu_0 \geq 0, j = 1, \ldots, n \\ \omega^T x_0 = 1 \\ \omega_i \leq \omega_j, 1 \leq i, j \leq n \\ \mu_i \leq \mu_j, 1 \leq i, j \leq n \\ \omega \geq 0, \mu \geq 0, \mu_0\end{cases} \tag{7}$$

无限制

2.3 Efficiency Analysis

(1) If $\theta_1^* = 1$ and $S_i^{-*} = S_r^{+*} = 0$, the DMU is both scale-effective and technique-effective;
(2) If $\theta_1^* < 1$ or $S_i^{-*} \neq 0, S_r^{+*} \neq 0$, the DMU is inefficient in scale efficiency or technical efficiency;
(3) For the $\theta_1^* < 1$ or $S_i^{-*} \neq 0, S_r^{+*} \neq 0$, the productive efficiency = θ_1, the technical efficiency = θ_2, and scale efficiency [4] = $\frac{\theta_1}{\theta_2}$;
(4) If the DMU is inefficient, we can provide optimization suggestions for the alternatives with the premise of unchanged input or unchanged output [5]:

$$\begin{cases} x_0^{'} = (1 - \theta_1^*)x_0 + S^{-*} \\ y_0^{'} = S^{+*} \end{cases} \tag{8}$$

S_i^-, S_i^+ are slack variables during the solving of the dual programming to be estimated.

3 Case Study

In this case, we use the DEA model to evaluate different choices of a civil aircraft purchasing.

3.1 Variable Selection

With regards to the life cycle cost of the military airplane and the definition of the life cycle management of the EAM, we divided the life cycle of the civil aircraft into 3 stages, the asset Acquisition, the asset Operation and Maintenance and the asset Retirement Disposal.

Because the indirect cost of the civil aircraft differs from airline companies in a great sense, so we just take the direct cost here. See for Fig. 1.

Given the life cycle cost of the civil aircraft, we take the cost of Acquisition, the cost of Operation and Maintenance and the cost of Retirement Disposal as the input variables and the Wind Loading and Thrust-to-weight Ratio as output variables [6].

The cost of the Acquisition (CACQ) includes the cost of the demonstration and development and test and evaluation of the scheme (CRDTE) the manufacture (CMAN) and the profit of the manufacturer of the airplane (CPRO).

The cost of the Operation and Maintenance refers to the cost generated during the actual operating and maintaining of the civil aircraft.

The cost of the Retirement Disposal is the disposal cost after the airplane is retired (Table 1).

3.2 Modelling Results

In the process of the modeling, firstly, we set the weights of the input variables free and then, with the fuzzy clustering and importance calculation, we make the weights of the output variables constrained. Then, with the use of LINGO, we get the below results (Table 2).

We can make some optimizations for the inefficient DMUs as well referring to the Eq. 8.

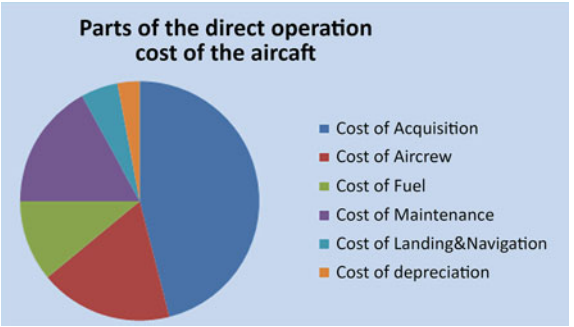


Fig. 1 Direct operation cost of the civil airplane

Table 1 Data comparison of different acquisitions [1]

Items		1	2	3
Input	CACQ (10^7 dollar)	7.29	6.01	6.61
	COPS (10^7 dollar)	8.78	6.14	7.34
	CDISP (million dollar)	4.97	3.76	4.31
Output	Wind loading (kg/m^2)	525.11	600.49	634.11
	Thrust-to-weight ratio	0.29	0.33	0.31

Data source The 22nd of the Handbook of the Aircraft Design: The design of the Technical Economy; The 5th of the Handbook of the Aircraft Design: The overall design of the civil aircraft

Table 2 Efficiency of the DMUs

Items	Productive efficiency	Technical efficiency	Scale efficiency
1	0.7244877	0.8244170	0.8787871
2	1	1	1
3	0.9601340	1	0.9601340

3.3 Data Analysis

From the previous results, we can know that first choice is the least efficient and the second is the most efficient. There is much work to be done for the aircraft 2 and 3 to be more competitive as shown in Table 3. In fact, the aircraft in Scheme 1 has been gradually eliminated because of its low competency. The aircraft in the second and third ones are once the most popular models but between their competitions, previously in the 13th Asian Aerospace, the second has won more and more advantage for its high reliability and low operation cost.

Table 3 Optimizations for the inefficient DMUs

Items	C_{ACQ} (10^7 dollar)	C_{OPS} (10^7 dollar)	C_{DISP} (10^6 dollar)
1	2.01	3.38	1.67
3	0.26	0.84	0.33

4 Conclusion

With the Object-weights-constrained DEA model and taking the life cycle management into consideration, we can get a clear overview of the engineering asset acquisition in aircraft industry. Also, with DEA model and LINGO, we can simplify the multi-objective programming and get the best choice through the comparison among the relative efficiency.

References

1. Chen HB, Xue YL, Han TXJ (2009) Trend and challenge of asset management for power engineering. East China Electric Power 01:81–85
2. Huang DX, YING KF, WU ZYJ (2003) Study and application on impersonal significance ordering with multi-factors. Ind Eng Manage 03:24–27
3. Zhu QJ (1994) Review and prospect of the DEA model. Syst Eng-Theory Pract 04:1–8
4. Wei QL M (2012) DEA and DEA network. In: 1st edn. Renmin University Press, Beijing
5. Wu YY, He XJJ (2006) The evaluation of Beijing sustainable development based on DEA model. Syst Eng-Theory Pract 03:117–123
6. Zhu DW, Wang TJ (2011) An engineering method to estimate aircraft performance criteria. Sci Technol Eng 11:31

Method of Measuring Mechanical Properties for Semi-Infinite Coating Materials

Guorong Song, Hongshi Liu, Zimu Li, Cunfu He and Bin Wu

Abstract At present coating materials are widely used in the field of aerospace, machinery, petroleum, chemical, nuclear power. The strength and failure analysis of coating is more and more important, at the same time the coating mechanical properties are critical for guiding the industrial electroplating. The elastic constant is one of the most important mechanical properties of parameters that are needed to be evaluated. The paper presents the measurement method of elastic constants for coating material based acoustic microscopy technology. The elastic constants of different thickness semi-infinite nickel coating materials are test by the $V(f, z)$ analytical method, it's a frequency domain method; which is to obtain the experimental dispersion curves. The mechanical properties of the nickel coating can be inversed by changing the longitudinal wave velocity and shear wave velocity of nickel coating to fit the theoretical dispersion curves with the experimental dispersion curves. The experimental results show the method is feasible, this study lays a foundation for evaluating mechanical properties for semi-infinite coating materials.

1 Introduction

Coating technology is regarded as a technology for the preparation of materials. It maintains the inherent feature of the base material, while also makes the surface of the material get characteristics such as anti-corrosion, antifriction and anti-oxidation. This technology is well-used in lots of fields. Coating strength and failure analysis become more and more important with the development of the coating technology and increasing requirements for modern equipment reliability, especially for testing mechanical properties of materials. Testing mechanical properties of metal coating is much more difficult compared with bulk materials. On one side,

G. Song (✉) · H. Liu · Z. Li · C. He · B. Wu
College of Mechanical Engineering and Applied Electronics Technology,
Beijing University of Technology, Beijing, China
e-mail: grsong@bjut.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_65

substrate and the coating interface have a great influence on the coating test results; on the other side, the methods applied on the block are often not applicable to the coating, so we are eager to establish new methods to achieve mechanical properties by non-destructive testing technology [1–3]. In order to analyse mechanical properties of coating materials, the representative testing methods are nano-indentation, bulge testing, micro-beam bending testing, etc. Above all, the method to test ultrasonic propagation characteristics and wave velocity is regarded as a focus recently.

In this paper, the frequency-domain a $V(f, z)$ analytical method, which is a kind of the ultrasonic velocity methods, is applied to inverse the mechanical properties of coating materials [4]. The mechanical properties of 10 mm aluminum substrate with 15, 35, 55, 60, 95 μm nickel coating layer semi-infinite samples are measured by the ultrasonic non-destructive testing system with a PVDF line-focus transducer developed by ourselves. The measurement result is quite satisfactory with high precision, and this system is reliable to obtain characterization and measurement of the coating semi-infinite material mechanical properties [5, 6].

2 Measurement Method

The measurement experiment for acoustic characteristics of nickel coating semi-infinite materials is based on acoustic microscope technology, using the ultrasonic non-destructive testing system for limited-size samples with a PVDF line-focus transducer developed by ourselves to complete defocus measurement [7]. The ultrasonic waves are excited and received by the continuously defocus short steps between line-focus transducer and the tested sample in order to get different echo signals at different positions.

The test dispersion curves can be got by two-dimensional Fourier transforming the received signals with $V(f, z)$ analytical method [3]. Firstly, the $V(f, z)$ oscillating curves can be obtained by Time-domain Fourier transform from the $V(t, z)$ echo signals, then, the $V(f, 1/z)$ oscillating curves can be got by Spatial-domain Fourier transform. Lastly, the period of defocusing distance Δz is the reciprocal of frequency peak, which is extracted from the $V(f, 1/z)$ oscillating curve. Take them into the formula:

$$V_R = V_W \left[1 - \left(1 - \frac{V_W}{2f\Delta z} \right)^2 \right]^{-\frac{1}{2}} \tag{1}$$

The test dispersion curves can be got, and V_W is the velocity of ultrasonic waves in the water.

A PVDF line-focus ultrasonic transducer developed by ourselves is applied, its parameters is: 20 mm focal distance, 5 MHz center frequency, 150° aperture angle. Figure 1 shows the PVDF line-focus ultrasonic transducer. Fixing the transducer on

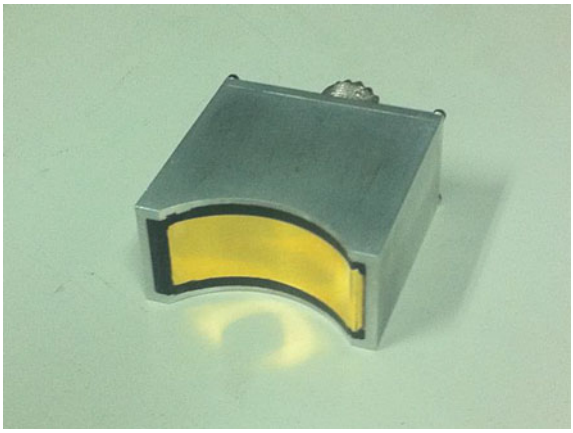


Fig. 1 PVDF line-focus ultrasonic transducer

Table 1 Dimension parameters of semi-infinite sample

Number of samples	Thickness of substrate (mm)	Thickness of nickel coating layer (μm)
1	10	15
2	10	35
3	10	55
4	10	60
5	10	95

the Z-axis of the Four-axis moving framework, this is an important part of ultrasonic non-destructive testing system. The transducer can automatically defocus with precise positioning by short steps (that is step = 10 μm).

There is a water tank with measured sample, it is put on a Three-axis adjust seat to keep PVDF line-focus transducer vertical with the surface of the measured sample [8, 9]. Nickel layers is widely used in functional coatings, the dimension parameters in this paper as shown in Table 1, All are semi-infinite samples for aluminum substrate with nickel coating layer.

When testing 10 mm aluminum substrate with 35 μm nickel layer semi-infinite sample, the defocus distance is 4 mm and step interval is 0.01 mm. So 401 groups echo signals can be got. The Time-domain $V(t, z)$ echo curves are showed as Fig. 2.

$V(f, z)$ analytic method is used to get the oscillating period by two-dimensional Fourier transform, because we can just get the oscillating curves from one-dimensional Fourier transform as in Fig. 3. Figure 3 is a frequency-domain waveform based on the composed oscillating curves of different frequency, each oscillating curve has a period at its optimum oscillation place [10].

The reciprocal of frequency peak can be regard as the period on z, so the defocus distance Δz is the reciprocal of the frequency peak of the acquired $V(f, 1/z)$ oscillating curves(shown in Fig. 4) after two-dimensional Fourier transform on defocus position z. Then, put them into the wave velocity formula to acquire the

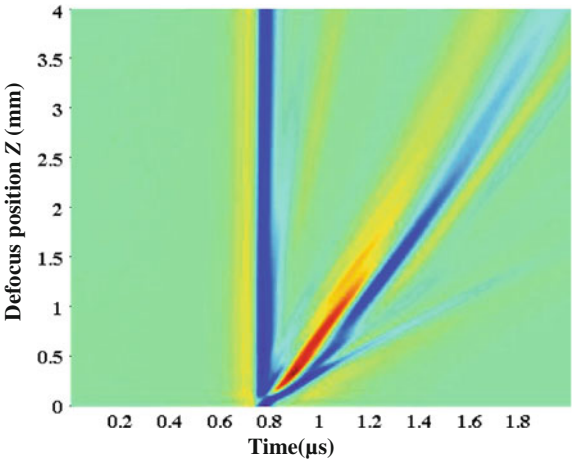


Fig. 2 Time-domain waveform of nickel coating layer (thickness 35 μm)

frequency dispersion curves of 10 mm aluminum substrate with 35 μm nickel coating layer semi-infinite sample, as showed in Fig. 5.

The velocity changes with the frequency, as its wavelength changes with frequency. The wavelength decreases accompany with the frequency increases, so the wave will penetrate to the substrate at the beginning, then to the coating layer. Figure 5 shows the relationship between frequency and velocity, an inflexion appears at 10 MH. This inflexion generate at the boundary of the substrate and coating layer, which indicates neither the wave velocity of two parts. It could be a method to get the thickness of coating layer by tracing the inflexion.

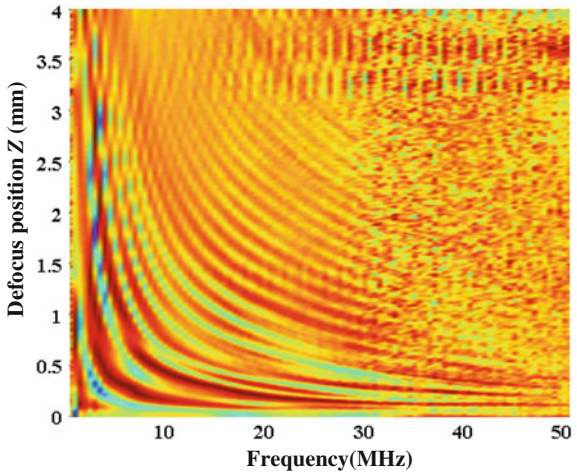


Fig. 3 Frequency-domain waveform of nickel coating layer (thickness 35 μm)

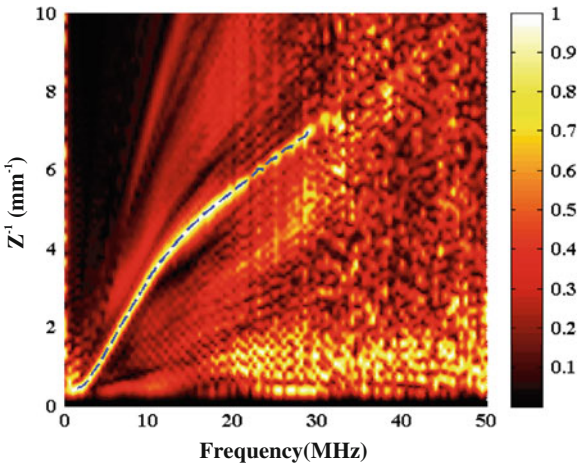


Fig. 4 $V(f, 1/z)$ oscillating waveform of nickel coating layer (thickness 35 μm)

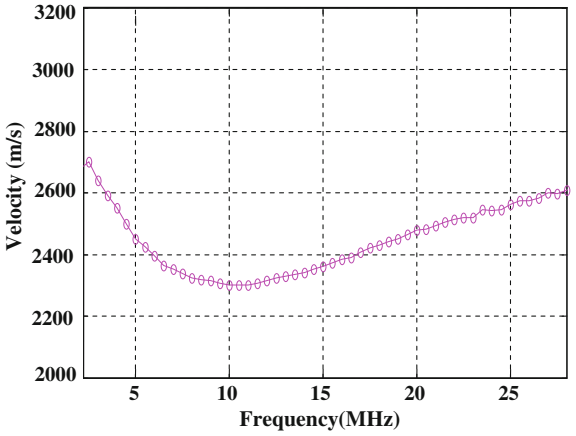


Fig. 5 Frequency dispersion curve of nickel coating layer (thickness 35 μm)

3 Measurement Result

The theory acoustic dispersion curve will be acquired by changing the longitudinal wave velocity and shear wave velocity, which involve acoustic wave propagation characteristics and theory modelling in the coating material. If the substrate density ρ, longitudinal wave velocity C_L, shear wave velocity C_T and nickel coating layer density ρ' are all given out, the theoretical values of nickel coating layer longitudinal wave velocity C'_L, shear wave velocity C'_T and its thickness d are those parameters make theory acoustic dispersion curve match test acoustic dispersion curve the best.

When inverse elastic constants of nickel coating layer semi-infinite materials, coating layer longitudinal wave velocity C_L range 5610–5810 m/s, shear wave

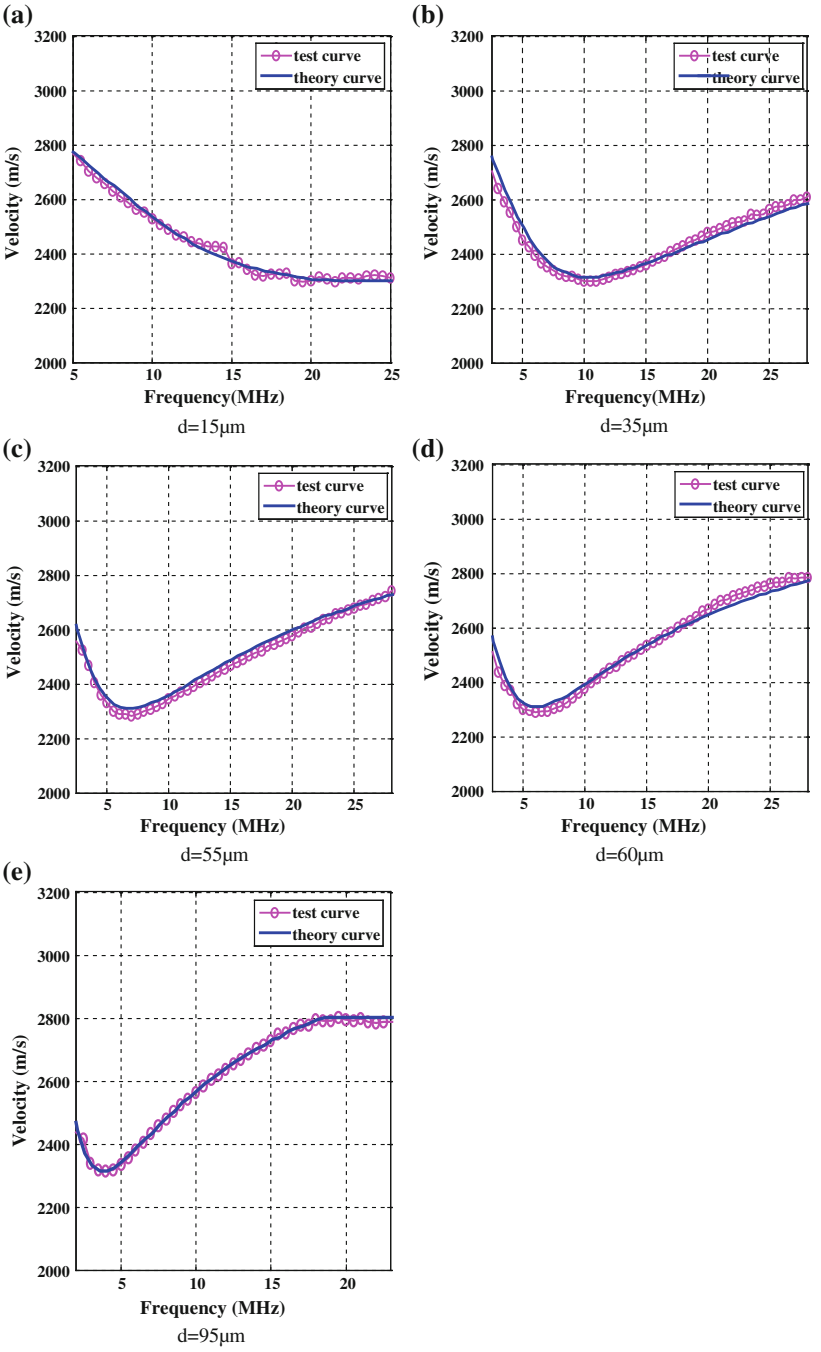


Fig. 6 The frequency dispersion curves of 5 different thicknesses coating layers (**a**)–(**e**). **a** d = 15 μm, **b** d = 35 μm, **c** d = 55 μm, **d** d = 60 μm, **e** d = 95 μm

Table 2 Comparison of results

Number of samples	Thickness of nickel coating layer μm	Longitudinal wave velocity C_L m/s	Shear wave velocity C_T m/s	Poisson's ratio υ	Young modulus E GPa
1	15	5830	2830	0.346	178.8
2	35	5830	2900	0.334	187.6
3	55	5820	2910	0.333	187.7
4	60	5810	2980	0.321	194.5
5	95	5800	3040	0.311	200.8

velocity range C_T 2930–3080 m/s can be used as a reference. The substrate density ρ is 2740 g/cm^3, and nickel coating layer density ρ' is 8300 g/cm^3. After several time testings, the 15, 35, 55, 60, 95 μm thickness fitting curves of coating layer are showed in Fig. 6a–e. The solid line is theory curve from changing parameters, while the dotted lines is our test curves obtained by $V(f, z)$ analytic method.

According to elastic mechanics theory, the mechanical properties are related to the velocity of the acoustic waves, when the velocity is given, as the following formula:

$$\upsilon = \frac{0.5C_L^2 - C_T^2}{C_L^2 - C_T^2} \tag{2}$$

$$E = \frac{C_L^2 \cdot \rho(1+\upsilon)(1-2\upsilon)}{1-\upsilon} \tag{3}$$

Poisson's ratio υ, Young modulus E can be obtained. ρ is the coating material density.

The coating layer thickness d can be directly obtained, other theoretical values are took into formula (2) and (3) to obtain Poisson's ratio υ, Young modulus E. Table 2 list the thickness of the test nickel coating layer, theoretical values for longitudinal wave velocity C_L, shear wave velocity C_T and Poisson's ratio υ, Young modulus E obtained by the measurement.

Figure 6 shows five different thicknesses nickel coating layer dispersion curves; they all have an inflexion at different frequencies. The inflexion moves to lower frequency with the layer gets thicker. Because the transducer is limited below 30 MHz, curves for the coating layers thinner than 15 μm are difficult to fit out. A high frequency transducer is needed to practise for thin layers.

In the Table 2 Poisson's ratio υ has a trend of decline and Young modulus E has a tread of increase, while thickening the coating layer. It means the mechanical properties of the nickel coating layers semi-infinite samples are closer to nickel, with thickening their coating layers.

4 Conclusion

The ultrasonic nondestructive testing system can realize the automatic measurement of the coated specimen acoustic characteristics. The measurement of echo signal is accuracy, and signal to noise ratio is high. In this paper, five aluminum substrates with nickel coating layer semi-infinite samples are measured by the $V(f, z)$ analytical method. The experimental results indicate that fit theory curve with test curve by changing the theoretical values based on $V(f, z)$ analytical method can help to inverse the mechanical properties and thickness of the nickel coating layers for semi-infinite sample. This study also lays a foundation for evaluating mechanical properties for semi-infinite coating materials.

Acknowledgements The work presented in this paper is supported by the National Natural Science Foundation of China (No. 11172014, 61271372, 51235001). The National Research Foundation for the Doctoral Program of Higher Education of China(No. 20091103110004). Beijing City Board of Education Science and Technology Plan(No. KM2010100050 34) and Research Foundation for the Doctoral Program of Beijing University of Technology.

References

1. Kulik VM, Lee I, Chun HH (2008) Physics of fluids. Wave properties of coating for skin friction reduction. Institute of Thermophysics, Russian Academy of Sciences, Novosibirsk
2. Reddy GM, Rao KS, Mohandas T (2009) Friction surfacing: novel technique for metal matrix composite coating on aluminium–silicon alloy. Surf Eng 25(1):25–30 (Defence Metallurgical Research Laboratory, Hyderabad, India)
3. Fu QG, Li H. J, Shi XH, Li KZ, Sun GD (2005) Scripta materialia. Silicon carbide coating to protect carbon/carbon composites against oxidation. Laboratory of Superhigh Temperature Composites, Xian, pp 923–927
4. Zhang F, Krishnaswamy S, Fei D, Rebinsky DA, Feng B (2006) Thin solid films. Ultrasonic characterization of mechanical properties of Cr-and W-doped diamond-like carbon hard coatings. Northwestern University, Evanston, pp 250–258
5. Alleyne DN, Cawley P (1990) Ultrasonics symposium. A 2-dimensional fourier transform method for the quantitative measurement of lamb modes. Imperial College of Science, London, Technology and Medicine, England, pp 1143–1146
6. Lee, Y. C. (2001). Ultrasonics. Measurements of dispersion curves of leaky Lamb waves using a lens-less line-focus transducer. Department of Mechanical Engineering, National Cheng Kung University, Tainan, pp 297–306
7. Guorong S (2012) Insight. Develop high-precision ultrasonic microscopy measurement system and measure the surface wave velocities of (100) silicon wafer. College of Mechanical Engineering and Applied Electronics Technology, Beijing University of Technology, Beijing, pp 253–256
8. Kushibiki J, Horii K, Chubachi N (1983) Velocity measurement of multiple leaky waves on germanium by line-focus-beam acoustic microscope using FFT.19. Department of Electrical Engineering Faculty of Engineering Tohoku University, Sendai, pp 404–405
9. Dixon S, Lanyon B, Rowlands G (2006) Coating thickness and elastic modulus measurement using ultrasonic bulk wave resonance. Appl Phys Lett 88(14):141907 (Department of Physics, University of Warwick, United Kingdom)
10. Cunfu H. Yan L, Guorong S (2011) Journal of mechanical engineering. Design, fabrication of line-focus lens-less polyvinylidene fluoride transducers and application on measuring surface acoustic waves with V(f, z) analytical method. College of Mechanical Engineering and Applied Electronics Technology, Beijing University of Technology, Beijing, pp 1–7

The Design of a MRE-Based Nonlinear Broadband Energy Harvester

Peter W. Tse and M.L. Wang

Abstract In this article, a conceptual design with its architecture of a broadband, vibration-based, nonlinear energy harvester is reported. Its non-linear behavior and its functionality are presented. Compared to that provided by conventional linear beam type of energy harvesters, this nonlinear harvester can provide widen the resonance frequency ranges. Hence, it can collect more vibration energy generated at various dominant rotational frequencies of a rotary machine. A smart material, called Magneto Rheological Elastomer (MRE), was added to the usual beam structured energy harvester. Since MRE is one of the magnetic smart materials, of which their stiffness can be tuned by precisely controlling the applied magnetic field, the stiffness of the combined piezoelectric beam with MRE becomes adjustable. Because of the adjustable stiffness, the resonance frequency of the new beam type energy harvester can be adaptively changed to match with a particular dominant rotational frequency generated by the monitored machine so that maximum vibration energy can be harvested. Moreover, due to the nonlinearity of the new composite beam structure, the range of resonance frequency range can be widened to make it easier to adapt a slightly varying dominant rotational frequency due to the monitored machine has small speed variation. Besides the presentation of design and its with its architecture in the article, the simulated and experimental results of the new non-linear harvester are also reported here. From the comparison study of the bandwidth and the output power generated by the new nonlinear energy harvester against that generated from conventional harvester, the result shows that new non-linear harvester is functioning superior to that of the conventional harvesters.

Keywords Energy harvesting · Vibration analysis · Smart materials · Nonlinear analysis

P.W. Tse (✉) · M.L. Wang
The Smart Engineering Asset Management Laboratory (SEAM), Department of Systems Engineering and Engineering Management (SEEM), City University of Hong Kong, Tat Chee Ave, Hong Kong, P.R.C
e-mail: Peter.W.Tse@cityu.edu.hk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_66

1 Introduction

Energy is one of the important topics that attracted increasing attention from academics and industry. Vibration based micro-generator is a promising energy scavenging method in collecting harmful energy and providing power supply to the mechanical health monitoring sensors. MRE is made up of carbonyl iron particles (ISP. grade S-3700) that embedded in natural rubber thus whose dynamic characteristic is reversely controllable by varying the external magnetic field [1]. Existing MR elastomers can be divided into two classes, isotropic and anisotropic, and the magneto-elastic effect partly dependent on the portion of carbonyl iron particles in the mixtures which maximized at about 33 % [2]. Kallio et al. also studied the dynamic properties such as the young modulus and damping ratio of the MRE. Since it is a soft material and its main characteristic is the damping which determined that it is mostly used as damper for vibration control [3, 4].

Currently, most researches of energy harvesting are carried out by basing on the piezoelectric effect and mechanical resonance including linear and nonlinear energy harvesters. Reviewing the existing linear and nonlinear energy harvester could provide insight to the design of the MRE based energy harvester. Erturk et al. [5] proposed a linear harvester which was an L-shaped beam-mass structure which is more stable and productive than the traditional one when subject to the random vibration. Zhou et al. [6] introduced a 2 DOF (degree of freedom) vibration magnifier, the relay beam, to enhanced the vibration of the attached piezoelectric materials. Huang and Lin added a moving support to the traditional 1DOF energy harvester which made the effective length of the vibration beam is tunable and hence the tunable resonance frequency [7]. Soliman et al. composed the micro-generator based on electromagnetic induction instead of piezoelectric and developed it by introducing a stiffness tuning stopper which broadens its resonance band [8].

Stanton et al. analytically solved the nonlinearity that introduced by a bras with PZT-5H attached on both sides (Stanton, [9, 10]. Sebald et al. analytically and experimentally validate the Duffing oscillator which lightened the structural nonlinearity to energy harvesting [11]. Stanton el al. also proposed a resonance magnification methods to enhance the output of energy harvester analytically and experimentally validate the Duffing oscillator which lightened the structural nonlinearity to energy harvesting. Cottone et al. showed a Piezoelectric buckled beams for bi-stable energy harvesting and the resonance frequency could be tuned by the degree to which the beam was buckled [12]. The bulked configuration presents a superior power generation over a large interval of resistive load when compared to the un-bulked ones. Abdelkefi [13] analytically solving the structural nonlinearity of the double-layer energy harvester which quit similar to this article. But the global model is suitable for objects that with strong stiffness instead of the viscoelastic material such as MRE. Zhou and Wang studied the sandwiched beam with MRE core [4, 14]. But, the face-plates are nonconductive and no electromechanical coupling effect. Hu et al. introduced the sandwich beam to energy harvesting and

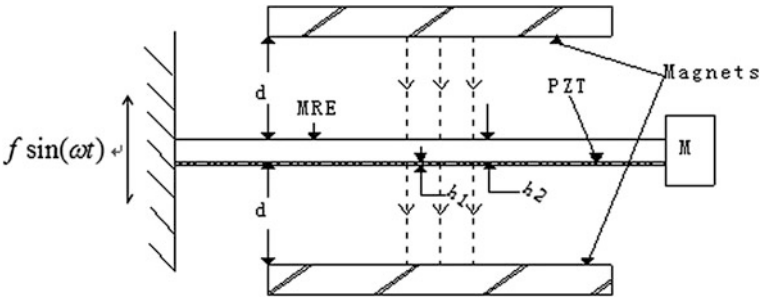


Fig. 1 Prototype of the MRE-based energy harvester

make the resonance frequency tunable as well as raised many more time by adding one preload to the end of the beam [15].

In this article, the MRE was used to composite the laminated beam for vibratory energy collection. The influences of parameters in the proposed designs are experimental study. In addition, by comparison of the bandwidth and the output power generated from the proposed generator with traditional ones, its superiority is showed in this article.

2 Proposed Design of Energy Harvester

The proposed design of broad band energy harvester is illustrated in Fig. 1. The design adopts adjustable magnet field strength by the changeable magnetic distance. As the distance of the magnet to MRE decrease, the magnet field strength increase and so do the storage modulus of the MRE accordingly. We assume the magnetic field density is constant in the gap of two magnets when they are at specific distance.

The deformation of MRE patch that attached to the piezoelectric materials mostly considered being the shear strain and regarded as viscoelastic material. The dynamic motion of MRE is mainly contributed by the complex modulus.

3 Experimental Studies

3.1 Experimental Settings

Figure 2 shows the experimental set up. The shaker was excited by a sweeping signal. The QP20 N is the piezoelectric material (PZT) based harvester. The MRE is the Magneto Rheological Elastomer, which is obtained from Ioniqa Technologies B.V. (Eindhoven, The Netherlands). Figure 3 shows six different configurations of

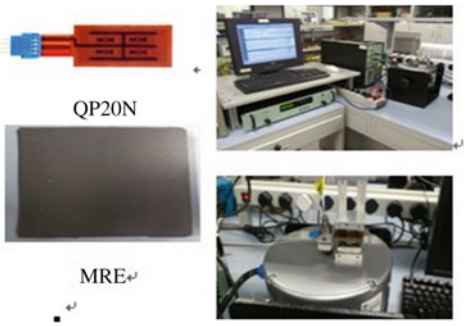


Fig. 2 The experimental setup

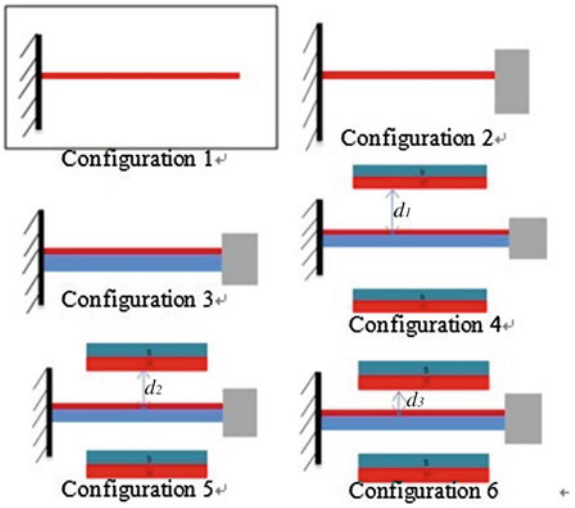


Fig. 3 Harvesters with 6 different configurations for comparison purpose

the tested harvesters. These configurations include the PZT material along (configuration 1), the PZT with a tip mass attached (configuration 2), the PZT and MRE with the tip mass (configuration 3), the PZT, MRE, the tip mass and two magnets located apart at a distance, *d*1 (configuration 4), and the same configuration but with different distances, *d*2 and *d*3 for configurations 5 and 6 respectively. Configuration 6 has the shortest distance, *d3*. The dimensions of harvester in different configurations are tabulated in Table 1. Their important parameters are tabulated in Table 2.

Table 1 Dimensions of the proposed design

Dimensions	Value	Unit
Thickness of the QP20 N h1	0.0006	m
Thickness of the MRE h2	0.003	m
Length of the QP20 N and MRE	0.0493	m
Width of the QP20 N and MRE	0.0254	m

Table 2 Important parameters

Parameters	Value	Unit
Young's modulus of QP20 N	55	GPa
Area moment of Inertia of QP20 N	4.57e-13	m^4
Density of QP20N	7700	kg/m^3
Distance from magnet to MRE *d*1	0.05	m
Distance from magnet to MRE *d*2	0.03	m
Distance from magnet to MRE *d*3	0.01	m
Magnetic strength B	Adjustable	Teslas
Mass of the tip mass (Al) M	1.5	g
Density of the MRE	2.5	g/ml
Magnetization of the MRE	160	KA/m

3.2 Experimental Validation of Intended Effectiveness

The curves in the Fig. 4 are the voltage response when the prototype excited in a same sweeping sine signal with 0.5 V amplitude and 1 MΩ resistance circuits. In order to compare the voltage response in different frequency range, we use the time

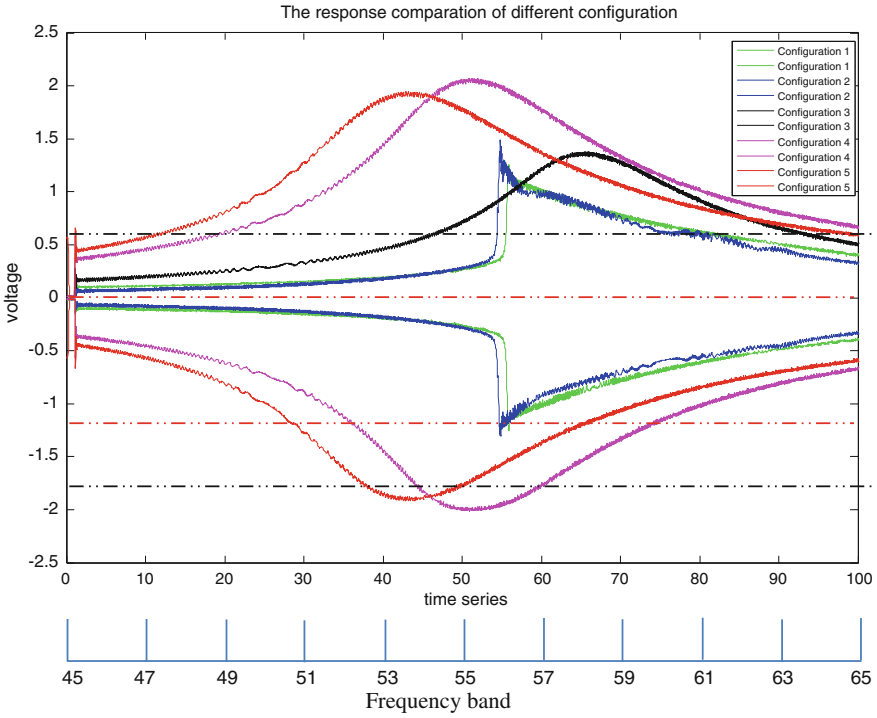


Fig. 4 The time-series voltage responses for different configurations

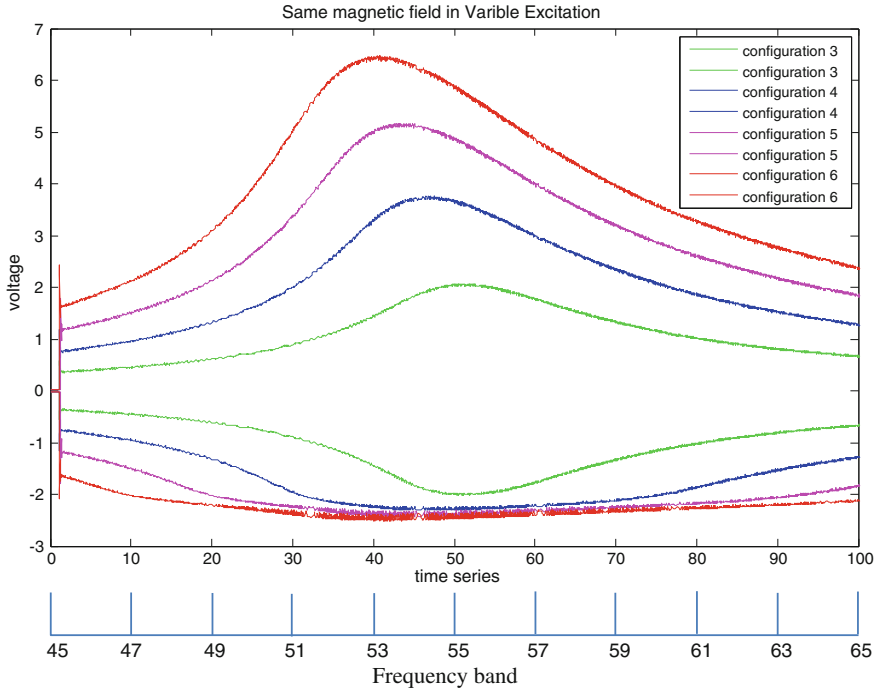


Fig. 5 The time-series voltage responses of different configurations under various excitations

as horizontal axis, and the sweeping rates are all at 0.2 Hz/s. The green curve is the voltage response of configuration 1 and blue for 2, black for 3, magenta for 4 and red for 5. As illustrated in Fig. 4, the green and blue curves present a sudden increase near their resonance frequency while in frequencies far from resonance the amplitude of voltage is considerable low. However, the amplitudes of the configuration 3–5, 1.35, 2.05 and 1.94 v, are much larger than the configuration 1–2, 1.25 and 1.45, which proves the proposed prototype is quite effective in increasing the energy collection.

Though the dynamic nonlinearity of the QP20 N could slight increase its resonance band, the band still quit narrow. As shown in Fig. 4, the areas enclosed by green, blue curve and red dash straight lines are about 4.9 and 6 VHz, while the areas enclosed by green, blue curve and black dash straight lines are about 0.5 and 0.04 VHz. So adding the tip mass to the end of QP20 N not only increases the response amplitude but also increases the resonance band. In this article, the resonance band means the frequency range adjacent to the resonance frequency point where the voltage response is relatively higher than the excitation frequency.

Configurations 3–5 are the prototype vibrating under various magnetic fields. The areas enclosed by the corresponding curve and red and blank straight line area about 9.89, 30, 28.35, 1.45, 9.6, 8.83 VHz. In the whole frequency band illustrated in Fig. 4, the total band energies (areas enclosed with the axis) of the different

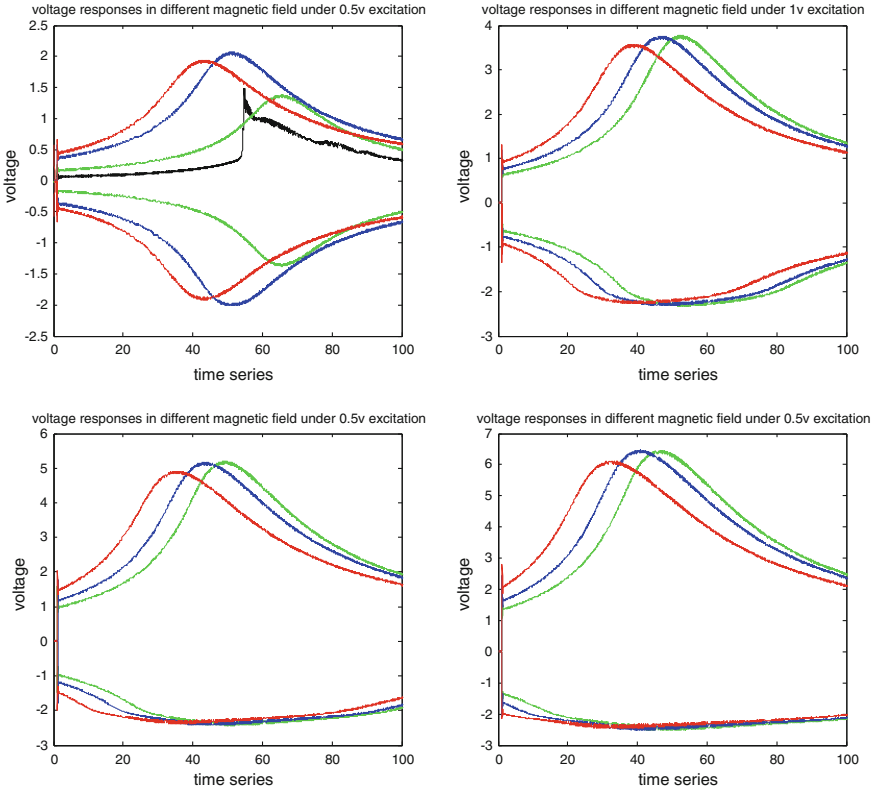


Fig. 6 Voltage responses of different configurations with same excitation but various magnetic fields

configurations are 14.9, 16, 19.89, 40, 38 VHz. Compared with the resonance band of configurations 1 and 2, we can see there is a significant enhancement in harvested energy.

3.3 Parameters Studies

When the prototype excited at different vibration amplitude, the resonance frequency varies as presented in Fig. 5. When the excitation becomes larger, the stain in the MRE increases accordingly. This leads to a decrease in its Young modulus. The complex modulus of the MRE can be formulated as defined in Eqs. 1 and 2 [16].

$$G(\omega) = G^R(\omega) + jG^I(\omega) = G^R(\omega)[1 + j\Delta(w)] \tag{1}$$

$$G^R = \alpha_0 + \alpha_1\omega + \alpha_2\omega^2, \Delta(\omega) = \beta_0 + \beta_1\omega + \beta_2\omega^2 \quad (2)$$

Since the shear modulus of MRE increases with the external magnetic field, so the resonance frequency of the under different magnetic field and vibration amplitude are studied as illustrated in Fig. 6. As the increase of magnetic field strength, amplitude of voltage increases to a maximum then decrease. As the excitation amplitude increases, the deformation of the MRE became larger which lead the steepening of the response curve. It should be noted that as the enhancement of excitation, the resonance frequency decrease which is clearly showed in the sub figures of Fig. 6. However, when the excitation is large enough, then the resonance frequency would not decrease any more. The asymmetric of the voltage response is due to the geometric nonlinear of the proposed structure.

4 Conclusion

The effectiveness of the new nonlinear harvest with added MRE for vibration energy harvesting has been experimentally validated. The design has largely broadened the resonance band which is also a new scale in measuring the bandwidth. The bandwidth of the new nonlinear harvester is twice larger than that from the traditional energy harvesters. The output of the electric energy density has also proved to be largely improved. The success of designing such novel nonlinear harvester with smart material provides a new perspective in the design of future energy harvesters.

Acknowledgments This article was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 122011) and a grant from City University of Hong Kong (Project No. 7008187). Magneto rheological Elastomer was kindly provided by Ioniqa Technologies B.V. (Eindhoven, The Netherlands).

References

1. Ginde JM, Nichols ME, Eliea LD, Tardiff IL (1999) . SPIE conference on smart materials technologies. Magnetorheological elastomers properties and applications.pdf
2. Kallio M. (2005) The elastic and damping properties of magnetorheological elastomers. VTT Publications, Espoo
3. Chen C, Liao W-H (2012) A self-sensing magnetorheological damper with power generation. Smart Mater Struct 21(2):025014
4. Zhou GY, Wang Q (2006) Study on the adjustable rigidity of magnetorheological-elastomer-based sandwich beams. Smart Mater Struct 15(1):59–74
5. Erturk A et al (2008) Modeling of piezoelectric energy harvesting from an L-shaped beam-mass structure with an application to UAVs. J Intell Mater Syst Struct 20(5):529–544
6. Zhou W et al (2012) An efficient vibration energy harvester with a multi-mode dynamic magnifier. Smart Mater Struct 21(1):015014

7. Huang S-C, Lin K-A (2012) A novel design of a map-tuning piezoelectric vibration energy harvester. Smart Mater Struct 21(8):085014
8. Soliman MSM et al (2008) A wideband vibration-based energy harvester. J Micromech Microeng 18(11):115021
9. Stanton SC et al (2010) Nonlinear piezoelectricity in electroelastic energy harvesters: modeling and experimental identification. J Appl Phys 108(7):074903
10. Stanton SC et al (2010) Resonant manifestation of intrinsic nonlinearity within electroelastic micropower generators. Appl Phys Lett 97(25):254101
11. Sebald G et al (2011) Simulation of a Duffing oscillator for broadband piezoelectric energy harvesting. Smart Mater Struct 20(7):075022
12. Cottone F et al (2012) Piezoelectric buckled beams for random vibration energy harvesting. Smart Mater Struct 21(3):035021
13. Abdelkefi A et al (2011) Global nonlinear distributed-parameter model of parametrically excited piezoelectric energy harvesters. Nonlinear Dyn 67(2):1147–1160
14. Zhou GY, Wang Q (2005) Magnetorheological elastomer-based smart sandwich beams with nonconductive skins. Smart Mater Struct 14(5):1001–1009
15. Hu Y et al (2007) A piezoelectric power harvester with adjustable frequency through axial preloads. Smart Mater Struct 16(5):1961–1966
16. Ying ZG, Ni YQ (2009) Micro-vibration response of a stochastically excited sandwich beam with a magnetorheological elastomer core and mass. Smart Mater Struct 18(9):095005

Feature Selection Approach Based on Physical Model of Transmission System in Rotary Aircraft for Fault Prognosis

Cheng Zhe, Hu Niao-Qing and Zhang Xin-Peng

Abstract A majority of the mishaps of rotary aircraft are caused by the faults in drive train which is composed of some complex rotary mechanical systems. Planetary gear sets are common mechanical components and are widely used to transmit power and change speed and/or direction in rotary aircrafts. Planetary gear sets are epicyclical gear drive that is more complex compared to ordinary gear train, so the features of planetary gear sets is quite different from traditional features and hard to extract. This research focuses on the physical-model-based approach to extract features for planetary gear set. Physical model will be established for planetary gear set with fault. Then, the features suitable for severity estimation is selected based on the simulation signals of physical models. After that, the tests with faults seeded are carried out, and the validation has a promising result.

Keywords Planetary gear sets · Physical model · Feature extraction · Feature selection · Prognostics

1 Introduction

The majority of mishaps in helicopters are caused by engine and drive train failures. To reduce these mechanically induced failures and excessive maintenance, it is vital to accurately identify and diagnose the developing faults in the mechanical system. Planetary gearbox is a common mechanical component and is widely used to transmit power and change speed and/or direction in rotary wing aircraft. One of the most common causes of planetary gear set failure is tooth fatigue crack of the sun gear due to excessive stress conditions. This failure causes progressive damage to

C. Zhe (✉) · H. Niao-Qing · Z. Xin-Peng
Laboratory of Science and Technology on Integrated Logistics Support, College of Mechatronics Engineering and Automation, National University of Defense Technology, Changsha 410073, Hunan, China
e-mail: chengzhe@nudt.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_67

gear teeth and ultimately leads to the complete failure of the planetary gear set. This fault is particularly challenging as it is located deep inside the main transmission, suggesting it would be difficult to detect earlier. As a result, the feature extraction for damage detection and severity estimation of planetary gear set is a challenge in the health management of helicopter.

In this paper, a physical-model-based feature selection approach for planetary gear set is proposed. Physical model will be established for planetary gear set with fault. Then, the features suitable for severity estimation is selected based on the simulation signals of physical models. After that, the tests with faults seeded are carried out to validate the features extracted above.

2 Physical Model of Planetary Gear Set with Defect

Although the modelling of healthy gear systems nowadays is extensively carried out, failure modelling is still the subject of many research papers. The finite elements method is the most frequently used technique to assess gear tooth failures by their meshing stiffness reduction, but it requires mesh refinements and then much computation time in certain applications [1–4]. Thus, this research is based on the analytical method, which focuses on the tooth stiffness reduction due to damage by considering qualitative proportional reduction [5–7].

2.1 Physical Model of Healthy Planetary Gear Set

A lumped parameter, pure torsional dynamical formulation is employed to develop the physical model of the 2 K-H planetary gear set. $K_{\mathrm{sp}i}$ denotes mesh stiffness between sun-gear and planet gear; $K_{\mathrm{rp}i}$ denotes mesh stiffness between planet gear and ring gear; $C_{\mathrm{sp}i}$ denotes mesh damping between sun-gear and planet gear; $C_{\mathrm{rp}i}$ denotes mesh damping between planet gear and ring gear.θ_{s}, $\theta_{\mathrm{p}i}$ and θ_{c} denote the rotation angle of sun-gear, planet gear and carrier respectively. T_{D}, and T_{L} denote driving torque and loading torque respectively. The subscripts s, r, pi and c denote sun-gear, ring gear, the ith planet gear and carrier. By ignoring mesh errors and defining internal meshing side clearance and external meshing side clearance as $2b_{\mathrm{sp}i}$ and $2b_{\mathrm{rp}i}$, respectively, the adhesive engaging force D and the elastic engaging force P are represented as:

$$\begin{cases} D_{\mathrm{sp}i} = C_{\mathrm{sp}i}(\dot{\theta}_{\mathrm{s}} r_{\mathrm{bs}} - \dot{\theta}_{\mathrm{p}i} r_{\mathrm{bp}i} - \dot{\theta}_{\mathrm{c}} r_{\mathrm{c}} \cos\alpha) \\ D_{\mathrm{rp}i} = C_{\mathrm{rp}i}(\dot{\theta}_{\mathrm{p}i} r_{\mathrm{bp}i} - \dot{\theta}_{\mathrm{c}} r_{\mathrm{c}} \cos\alpha) \\ P_{\mathrm{sp}i} = K_{\mathrm{sp}i}(t) f(\theta_{\mathrm{s}} r_{\mathrm{bs}} - \theta_{\mathrm{p}i} r_{\mathrm{bp}i} - \theta_{\mathrm{c}} r_{\mathrm{c}} \cos\alpha, b_{\mathrm{sp}i}) \\ P_{\mathrm{rp}i} = K_{\mathrm{rp}i}(t) f(\theta_{\mathrm{p}i} r_{\mathrm{bp}i} - \theta_{\mathrm{c}} r_{\mathrm{c}} \cos\alpha, b_{\mathrm{rp}i}) \end{cases} \tag{1}$$

where $K(t)$ is time varying mesh stiffness, C is mesh damping constant, and r_b is radius of basic circle. The nonlinear clearance function $f(x, b)$ is defined by:

$$f(x,b) = \begin{cases} x-b & (x>b) \\ 0 & (-b \le x \le b) \\ x+b & (x<-b) \end{cases} \tag{2}$$

where b is the clearance constant.

Dynamical differential equations of the 2 K-H planetary gear set could be deduced from Lagrange equations:

$$\begin{cases} I_s\ddot{\theta}_s + \sum_{i=1}^{N}(D_{spi}+P_{spi})r_{bs} = T_D \\ I_{pi}\ddot{\theta}_{pi} - (D_{spi}-D_{rpi}+P_{spi}-P_{rpi})r_{bpi} = 0 \\ \left(I_c + \sum_{i=1}^{N} m_{pi}r_c^2\right)\ddot{\theta}_c - \sum_{i=1}^{N}(D_{spi}+D_{rpi}+P_{spi}+P_{rpi})r_c\cos\alpha = -T_L \end{cases} \tag{3}$$

where m is mass, I is rotational inertia of the subcomponents (for planet gear, sun gear and carrier, $I = \frac{1}{2}mr_b^2$), α is mesh angle and N is the number of planet gear.

The equations above are positive semi-definite, nonlinear equations, which have N + 2 degrees-of-freedom (dof's), with angle displacements of rigid bodies in a coordinated system. For translating angle displacements of rigid bodies into relative linear displacement, the relative displacements between sun-gear and planet gear x_{spi}, and the relative displacement between sun-gear and carrier x_{sc} are defined as:

$$\begin{cases} x_{spi} = \theta_s r_{bs} - \theta_{pi} r_{bpi} - \theta_c r_c \cos\alpha \\ x_{sc} = \theta_s r_{bs} - 2\theta_c r_c \cos\alpha \end{cases} \tag{4}$$

After this, equivalent mass is supposed as M, for planet gear and sun gear, $M = I/r_b^2$, for carrier, $M_c = \frac{I_c}{r_{bc}^2} + \sum_{i=1}^{N}\frac{m_{pi}}{\cos^2\alpha}$.

Substituting Eqs. (1) and (3) into Eq. (4) and simplifying the equations, then the dynamic model of planetary gear sets is represented as:

$$\begin{cases} \ddot{x}_{spi} + \sum_{i=1}^{N}\left[\left(\frac{1}{M_s}+\frac{1}{M_c}\right)C_{spi}\dot{x}_{spi} + \frac{1}{M_c}C_{rpi}\left(\dot{x}_{sc}-\dot{x}_{spi}\right)\right] + \frac{1}{M_{pi}}\left[C_{spi}\dot{x}_{spi} - C_{rpi}\left(\dot{x}_{sc}-\dot{x}_{spi}\right)\right] \\ + \sum_{i=1}^{N}\left[\left(\frac{1}{M_s}+\frac{1}{M_c}\right)K_{spi}(t)f\left(x_{spi},b_{spi}\right) + \frac{1}{M_c}K_{rpi}(t)f\left(x_{sc}-x_{spi},b_{rpi}\right)\right] \\ + \frac{1}{M_{pi}}\left[K_{spi}(t)f\left(x_{spi},b_{spi}\right) - K_{rpi}(t)f\left(x_{sc}-x_{spi},b_{rpi}\right)\right] = \frac{T_D}{M_s r_{bs}} + \frac{T_L}{M_c r_c\cos\alpha}; \\ \ddot{x}_{sc} + \left(\frac{1}{M_s}+\frac{2}{M_c}\right)\sum_{i=1}^{N}C_{spi}\dot{x}_{spi} + \frac{2}{M_c}\sum_{i=1}^{N}C_{rpi}\left(\dot{x}_{sc}-\dot{x}_{spi}\right) + \left(\frac{1}{M_s}+\frac{2}{M_c}\right)\sum_{i=1}^{N}K_{spi}(t)f\left(x_{spi},b_{spi}\right) \\ + \frac{2}{M_c}\sum_{i=1}^{N}\left[K_{rpi}(t)f\left(x_{sc}-x_{spi},b_{rpi}\right)\right] = \frac{T_D}{M_s r_{bs}} + \frac{2T_L}{M_c r_c\cos\alpha}. \end{cases} \tag{5}$$

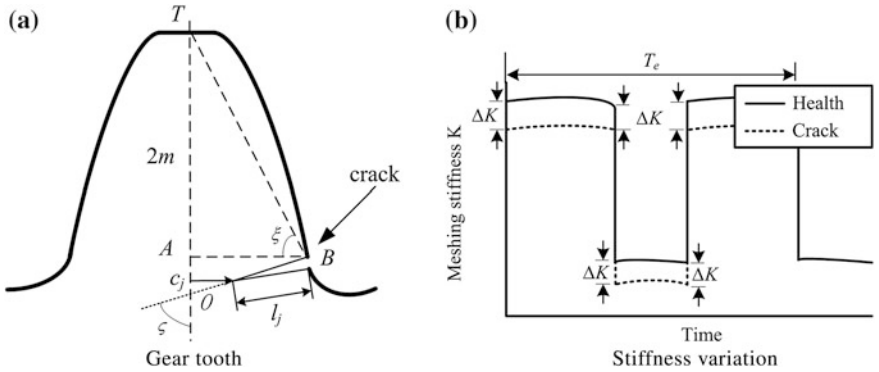


Fig. 1 Gear-mesh stiffness variation of gear pair with tooth crack

2.2 Physical Model of Planetary Gear Set with Defect

In this section, tooth fatigue crack in sun gear of planetary gear set is modelled. The common causes of this damage include cyclic stressing of the gear tooth material beyond its endurance fatigue limit. Bending fatigue crack starts in the root section and progresses until the tooth or part of it breaks off. Fatigue crack always occur in the usual tooth root fillet section, as can be seen in Fig. 1a.

To simplify the model of damage, at each section of the tooth, the shape of tooth root crack is approximated with straight line which is defined by the length l and the direction angle ς of the crack, as shown in Fig. 1. The tooth crack level is defined as s, which is determined by l and ς, $s = s(l_j, \varsigma)$. For simplifying the dynamical model of planetary gear set, the crack size and its threshold value are defined as c_j and c_f, $c_j = 2m\cot\xi - l_j\sin\varsigma$, $\tilde{c}_j = c_j/c_f$, $j = 1, 2, 3\ldots$. And then $s_j = s(l_j, \varsigma) = s(\tilde{c}_j)$, $s_j \in [0\ \%,\ 100\ \%]$. Referred to Literature [5, 6], the gear mesh stiffness of the gear pair, which is composed of sun-gear and planet-gear, is calculated by taking into account the geometric change due to the tooth crack. The details on the exact relationship between the crack size and the stiffness reduction can be referred to [4].

The gear mesh stiffness variation caused by sun gear tooth crack is defined as $\Delta K(f_d, t)$, so the time varying mesh stiffness with sun gear tooth crack can be expressed as

$$K_{spi}(t) = K_0(f_0, t) + \Delta K(f_d, t) \tag{6}$$

where $K_0(f_0, t)$ is the mesh stiffness of gear pair in healthy case, f_0 is the mesh frequency, and f_d is the meshing frequency of sun gear tooth with damage. f_d is the function of the rotary frequency of sun gear f_s, the rotary frequency of carrier f_c and the number of planet gears N, $f_d = N(f_s - f_c)$. The time vary mesh stiffness is illustrated in Fig. 1b. The dynamical model of 2 K-H planetary gear set with sun gear tooth crack is acquired by substituting Eq. (6) into Eq. (5).

Table 1 Parameter value in the models

Parameter (Unit)	Value	Parameter (Unit)	Value
Modulus (mm)	2.5	r_c (m)	0.108
Tooth number of sun gear	28	Driving torque (N•m)	100
Tooth number of planet gear	32	Loading Torque (N•m)	200
Tooth number of ring gear	92	K_{spi} (N/m)	1.806×10^9
Number of planet gear	4	K_{rpi} (N/m)	2.212×10^9
Tooth width (mm)	12	C_{spi} (N•s/m)	4.474×10^3
Pressure angle (Deg)	20	C_{rpi} (N•s/m)	5.764×10^3
m_s (kg)	1.144	b_{spi}, b_{rpi} (m)	0.0001
m_{pi} (kg)	1.757	I_{pi} (m)	0.0016
m_r (kg)	3.016	I_s (m)	0.0011
r_{bs} (m)	0.046	I_c (m)	0.032
r_{bpi} (m)	0.057	Material	40Cr

2.3 Simulation of the Physical Models

The parameters in the physical models above should be set up according to Table 1, for the faulty case, s is varied as [0:5:100 %]. The simulation is executed in Matlab, and four-order Runge-Kutta method is selected for the solution of models. The duration size and the sampling frequency in solving the equations are set up as 10 s and 10 k respectively.

Figure 2 shows the dynamic responses in the frequency domain for a healthy planetary gearbox (s = 0 %) and 3 different levels of crack (s = 20, 40, 80 %). The healthy model is characterized by the dominance of the gear mesh frequency, denoted by 1X and its harmonics of 2X ~ 6X, and etc.

For the damage seeded models, amplitude modulation in time domain of the gear mesh signal can be clearly observed. As a consequence, many new frequency components appear at the left side of the dominant frequency (near 1/2X). The amplitude of new frequency components increases with the growth of crack level, but the amplitude of dominant frequency and its harmonics decreases at the same time.

2.4 Validation of the Physical Models

The data sets of health condition and damage seeded condition, which are obtained from test rig of planetary gear set, are utilized to validate the physical model above. As the dominant frequency (1X) and its harmonics (2X ~ 7X) are the main components in frequency domain, therefore these components are normalized, the results of which are listed in Table 2. The frequency vector is created based on normalized frequency components (1X ~ 7X). After that, for validating the

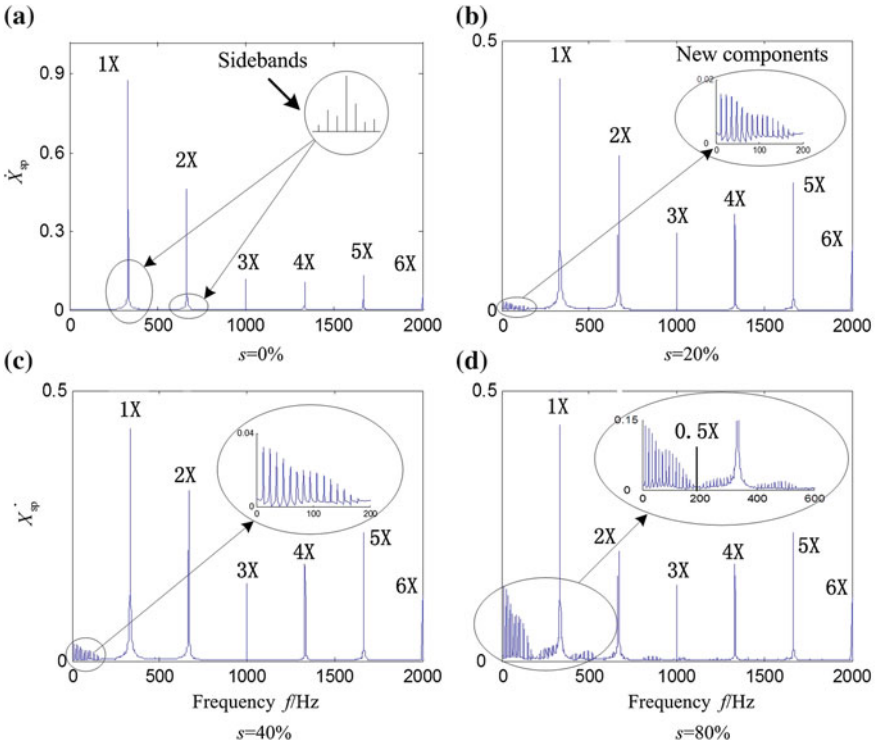


Fig. 2 Dynamic responses of physical models in the frequency domain

physical models above, two- sample Z test, which is used to analyze the statistical dependence of two data sets quantitatively [8, 9], is carried out between two frequency vectors of test data and simulation data. The equation of two-sample Z-test is as bellows:

$$Z = \frac{|\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2|}{\sqrt{\frac{S_{\mathbf{X}_1}^2}{n_1} + \frac{S_{\mathbf{X}_2}^2}{n_2}}} \tag{7}$$

where $\mathbf{X}_1$and $\mathbf{X}_2$ are the frequency vector of simulation data and test data, $S_{\mathbf{X}}$ and $\bar{\mathbf{X}}$ are the standard deviation and mean of $\mathbf{X}$, respectively. And n is the element number of frequency vector. This algorithm is used here to calculate the statistical difference of simulation signal and test data, the results are shown in Table 2. As can be seen in Table 2, the Z-test results are 0.2389 and 0.2586 for the health condition and the damage seeded condition, which show that simulation data and test data have no obvious difference on statistic, so the dynamical models above are validated to be effective.

Table 2 Validation results of the physical models

condition	model	1X	2X	3X	4X	5X	6X	7X	Z
Health	Math model	1.00	0.5611	0.1201	0.1475	0.1965	0.02969	0.02969	0.2389
	Test data	1.00	0.4469	0.4881	0.2041	0.1547	0.06268	0.02467	
Damage seeded	Math model	1.00	0.5602	0.1201	0.1485	0.1967	0.10030	0.02928	0.2586
	Test data	1.00	0.4207	0.4998	0.2485	0.2588	0.02150	0.01967	

3 Feature Selection Based on Simulation Signal

From our literature review [8–13], 27 features have been used in different cases of gearbox condition monitoring as lised in Table 3. In this study they are all explored to obtain an optimal subset for the detection and estimation of damage level in planetary gear sets and assigned with serial numbers. The simulation data sets generated by the dynamical models are used to calculate all these feature parameters in Table 3.

The approach of damage level estimation consists of two stages: damage detection and damage level identification. An optimal feature suitable for damage level estimation should have two merits: the first one is sensitivity, which means that the feature has a wider classification distance; the next one is relational, which means the feature is closely related to the fault. The target features will be selected based on the simulation data sets of the physical models developed above.

3.1 Sensitivity Analysis

In this research, the algorithm of two-sample *Z*-test is applied as the sensitivity analysis algorithm directly, and then it is modified to analyze the relational of the features. According to Eq. (7), the classification distance *Z* of the healthy samples

Table 3 Feature parameters for selection

Serial number	Feature parameter	Serial number	Feature parameter	Serial number	Feature parameter
1	Root mean squared (RMS)	10	FM0	19	M8A*
2	Crest factor (CF)	11	FM4	20	Mean frequency (MF)
3	Energy ratio (ER)	12	NA4	21	Frequency centre (FC)
4	Kurtosis	13	M6A	22	Root mean square frequency (RMSF)
5	Standard deviation	14	M8A	23	Standard deviation frequency (STDF)
6	Energy operator	15	NB4	24	Intra-revolution energy variance (IREV)
7	Absolute mean value	16	NA4*	25	Spectrum kurtosis (SK)
8	Clearance factor	17	NB4*	26	Local spectrum kurtosis (LSK)
9	Impulse factor	18	M6A*	27	NSR

and the fault seeded samples is calculated by two samples Z-test procedure. In the sensitivity analysis, $\mathbf{X}_1$ and $\mathbf{X}_2$ are the parameter vectors calculated based on healthy sample set and the fault seeded sample set respectively. And n is the sample number for each sample set. As the value of Z increases, the sensitivity of the feature parameter will grow, that means this feature parameter will have an obvious change while there is damage.

3.2 Relational Analysis

Whether a feature is a suitable indicator for crack level estimation is based on its performance to track crack propagation, which is called relational in this research. Only the feature curve is monotonic and close to the crack propagation curve, it can be considered that the feature has optimal relational. For quantitatively analyzing the feature performance on tracking crack evolution, the crack level vector is defined as a step curve.

In this section, two-sample Z-test is modified to make sure that the value of R can indicate the relational performance of feature parameter. Thus we exchange the numerator and the denominator in Eq. (7) each other to obtain Eq. (8). After that, the feature vector (real line) and damage level vector (imaginal line) are used as two samples respectively. The featuer vector is [a_1, a_2, a_3, a_4, a_5], and the damage level vector is [s_1/100, s_2/100, s_3/100, s_4/100, s_5/100]. Then the relational value R_k of the kth feature can be calculated following Eq. (8),

$$R_k = \frac{\sqrt{\frac{S_{\mathbf{X}_k}^2}{n_{\mathbf{x}_k}} + \frac{S_{\mathbf{V}}^2}{n_{\mathbf{V}}}}}{|\bar{\mathbf{X}}_k - \bar{\mathbf{V}}|}. \tag{8}$$

where $\mathbf{X}_k$ is the kth feature vector calculated based on damage propagation data, $\mathbf{V}$ is the damage level vector, $n_{\mathbf{x}_k}$ and $n_{\mathbf{V}}$ are the element number of $\mathbf{X}_k$ and $\mathbf{V}$, respectively. $\bar{\mathbf{X}}_k$ and $\bar{\mathbf{V}}$ are the mean of $\mathbf{X}_k$ and $\mathbf{V}$, respectively.

3.3 Feature Weighting

While the feature vector includes more than one element, feature weighting is necessary to fuse the estimation results of all parameters in feature vector and achieve more reliable results. In this research, the weighting is determined based on the relational performance of feature parameter. And also note that the weighting sum of all the feature parameters is equal to 1. So the parameters in the feature vector can be weighted as follows,

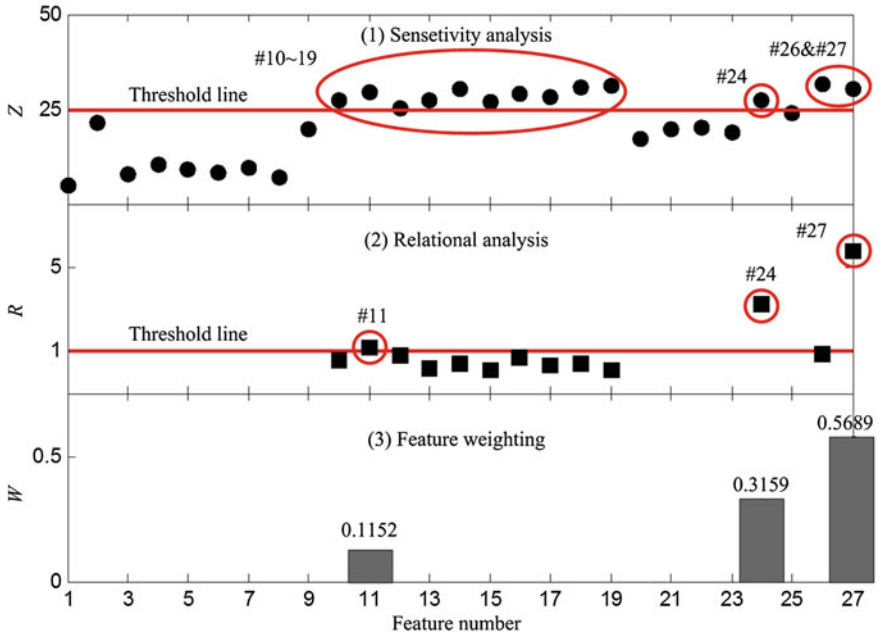


Fig. 3 Feature selection results

$$\begin{cases} W_1/R_1 = W_2/R_2 = \cdots = R_i/R_i = \cdots = R_n/R_n \\ W_1 + W_2 + \cdots + W_i + \cdots + W_n = 1 \end{cases} \tag{9}$$

3.4 Feature Selection

c weighting has been carried out with simulation data of dynamical models. The feature selection results show that only a few very important features are kept and used, as can be seen in Fig. 3. As there is no prior knowledge of the threshold setting, a median value is used as selection criteria. In this research, we have selected [25, 1] to be the threshold vector to illustrate the proposed methodology. As shown in Fig. 3, parameters #11, #24 and #27 have been selected to be the remaining features. These features are labelled as F1, F2 and F3 in this paper for convenience.

4 Validation with Test Data

A number of damage seeded experiments with different crack levels have been conducted on the test rig. The characteristics of the planetary gear set are given in Table 1. The vibration signal generated by the planetary gearbox was picked up by

four accelerometers bolted on the planetary gearbox casing and the electrical signal was transferred to the data acquisition system, which has a fore-charge-amplifier. The sampling frequency f_s is 10 kHz. The signal was low-pass filtered at 5 kHz through a 4th order Bessel type filter, in order to limit aliasing distortion and retain waveform integrity as much as possible. Data was stored for post processing to a PC. A data set has been acquired in all experiments corresponding to a time-history length of 10 s as shown in Fig. 4.

The results of damage level estimation are shown in Fig. 5. It can be seen that the result curves are corresponding to the related crack levels. To confirm the result precision, we check the test record and obtain the crack levels of the 4 test data sets above as s1 = 5 %, s2 = 10 %, s3 = 22.5 %, s4 = 35 %. The test records agree well with the results of crack level estimation.

Fig. 4 Configuration of sun gears with different damage seeded

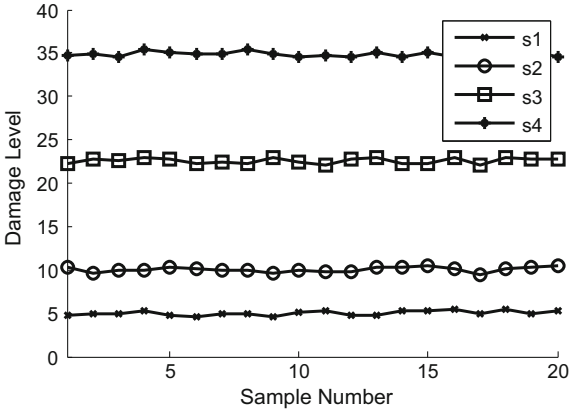


Fig. 5 Test validation results

5 Conclusion

This research focused on the physical-model-based approach to select features for planetary gear set. Some feature parameters were selected and weighted based on the simulation signals of physical models and some statistical algorithms. Finally, the tests with faults seeded are carried out to validate the performance of feature parameters extracted above.

Acknowledgments This research is supported by the National Natural Sciences Foundation of China (Grant Number 51205401) and Research Fund for the Doctoral Program of Higher Education of China (Grant Number 20124307120010). Valuable comments on the paper from anonymous reviewers are very much appreciated.

References

1. Pimsarn M, Kazerounian K (2002) Efficient evaluation of spur gear tooth mesh load using pseudo-interference stiffness estimation method. Mech Mach Th 37:769–786
2. Wang J (2003) Numerical and experimental analysis of spur gears in mesh. Ph.D., Curtin University of Technology
3. Ambarisha V, Parker R (2007) Nonlinear dynamics of planetary gears using analytical and finite element models. J Sound Vib 302:577–595
4. Lin J, Parker R (1999) Sensitivity of planetary gear natural frequencies and vibration on modes to model parameters. J Sound Vib 228:109–128
5. Chaari F, Fakhfakh T, Haddar M (2006) Dynamic analysis of a planetary gear failure caused by tooth pitting and cracking. J Fail An Prev 2:39–44
6. Chaari F, Baccar W (2008) Effect of spalling or tooth breakage on gearmesh stiffness and dynamic response of a one-stage spur gear transmission. Eur J Mech A/Solids 27:691–705
7. Hbaieb R, Chaari F, Fakhfakh T, Haddar M (2005) Influence of eccentricity, profile error and tooth pitting on helical planetary gear vibration. Mach Dyn Prob 29:5–32
8. Dempsey P, Lewicki D, Le D (2007) Investigation of current methods to identify helicopter gear health[C]. In: 2007 IEEE Aerospace conference, big sky, montana, march 3–10
9. Samuel P, Pines D (2005) A review of vibration-based techniques for helicopter transmission diagnostics. J Sound Vib 282:475–508
10. Saxena A, Wu (2005) B. Vachtsevanos G, A methodology for analyzing vibration data from planetary gear systems using complex morlet wavelets. In: 2005 American control conference, Portland, OR, USA, 8–10 June 2005
11. Wu B, Saxena A (2004) An approach to fault diagnosis of helicopter planetary gears[C]. In: 2004 IEEE autotestcon, 20–23, Sept 2004, pp 475–481
12. Barszcz T, Randall R (2009) Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine. Mech Syst Sig Proc 23:1352–1365
13. Cheng Z, Hu N, Qin G (2010) A new feature for monitoring of planetary gear sets based on physical model. In: Proceedings of the 23rd international congress on condition monitoring and diagnostic engineering management, June 28–July 2, 2010, Nara, Japan, pp 289–296

Machinery Fault Signal Reconstruction Using Time-Frequency Manifold

Xiangxiang Wang and Qingbo He

Abstract Machinery fault signals generally represent as periodic transient impulses, which often associate with important measurement information for machinery fault diagnosis. However, the existence of much background noise in practice makes it difficult to detect the transient impulses. Thus, it is very necessary to de-noise the measured signal and extract the intrinsic machinery fault signal for a reliable fault diagnosis. In this chapter, a novel de-noising method based on the time-frequency manifold (TFM) is proposed. This method mainly includes the following several steps. First, the phase space reconstruction (PSR) is employed to achieve a group of high-dimensional signals. For each dimensional signal, the short-time Fourier transform (STFT) is then conducted. Third, a suitable band carrying fault information is used for learning the TFM. Finally, the TFM is used to reconstruct the fault signal based on time-frequency synthesis and PSR synthesis. As the TFM has the merits of noise suppression and resolution enhancement to represent the inherent time-frequency structure, the reconstructed fault signal also has satisfactory de-noising effect, as well as good effect of inherent transient feature keeping. The proposed method has been employed to deal with a set of bearing data with rolling-element defect and outer-race defect, and the results show that the method is rather superior to two traditional methods in machinery fault signal de-noising.

1 Introduction

Signal de-noising has always been an important task in signal processing, and it is also increasingly significant in the field of electronic measurement and instrument. For machinery fault signals, such as vibration signals from defective rolling element bearings, they generally represent as periodic transient impulses due to the rotating

X. Wang · Q. He (✉)
Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei 230026, Anhui, People's Republic of China
e-mail: qbhe@ustc.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_68

nature. Research has shown that these periodic transient impulses often associate with important physical information related to the machine dynamics, so effective analysis of this kind of signals is the basis of machinery fault diagnosis. However, there always exists much background noise in collected signals in practice, which will corrupt the fault-induced transient impulses. Thus, it is always an important aim to de-noise the measured signal and extract the intrinsic machinery fault signal for a reliable fault diagnosis.

Generally, signal de-noising can be conducted in time domain, frequency domain, and time-frequency domain. The former two approaches can't take time and frequency information into account simultaneously, so the information of transient impulses will always be lost or the noise will not be removed completely. On the contrary, time-frequency representation can combine time and frequency information together, which can benefit de-noising with a synthetic consideration of both kinds of information [1]. Due to this merit, the time-frequency domain de-noising approaches have been widely developed [2–6]. Typical approaches are mainly based on the wavelet transformation (WT) and the short-time Fourier transform (STFT).

The WT has the characteristic of multi-resolution analysis, which is very suitable to detect a transient state anomaly that is embedded in normal signal. Hence, de-noising based on the WT is a hot spot of research. Wavelet threshold de-nosing method is very popular at present. However, there are also some problems. It is hard to select optimal wavelet basis for signal de-noising [2, 3] to avoid the loss of useful components in signal, and there is not a unitive and effective method to choose the threshold value [3–5].

The time-frequency analysis has the merit that it can intuitively represent the time-frequency information in a two-dimensional domain. The STFT threshold de-noising (also called spectrum subtraction) method is a typical one, especially for speech signal de-noising [6]. There are still some remained issues for this method, which makes that the de-noising effect is often not satisfactory. One of the most important issues is how to correctly distinguish noise in the time-frequency domain. This problem is hoped to be solved by the time-frequency manifold (TFM) technique, which is proposed by our group [7]. The TFM combines the benefits of time-frequency analysis in representing the non-stationary information and manifold learning in extracting intrinsic nonlinear structure of high-dimensional data, so it has the merits in noise suppression and resolution enhancement in the time-frequency domain. These merits benefit signal de-noising based on the time-frequency analysis approach.

This chapter intends to derive a novel signal de-noising method, which is to apply the TFM in noise reduction of measured noisy signals for a better machinery fault signal reconstruction. The basic idea of this method is to reconstruct the clear fault signal from the TFM of raw signal. As the TFM is a time-frequency structure with a high resolution for representing interesting impulse components and excellent suppression for the noise, theoretically the signals reconstructed from the TFM will have satisfactory de-noising effect. In the rest of the chapter, the signal reconstruction theory using TFM is introduced in Sect. 2 to derive the de-noising

method. Then, the effectiveness of the proposed method is verified by applications to a set of practical bearing signals in Sect. 3. Finally, conclusions are drawn in Sect. 4.

2 Signal Reconstruction Using TFM

The proposed signal de-noising method is motivated by the TFM's merits. As manifold learning can keep the nonlinear structure from high-dimensional data matrix, the TFM is a time-frequency structure with noise suppression, which can also represent the nature of original signal. This study introduces TFM to signal de-noising field. By combining with the technique of time-frequency synthesis and PSR synthesis, the de-noised signal can be constructed from the TFM. It aims to reduce background noise of signals effectively, and at the same time keep the essence of transient signals to the maximum extent. The following introduces the details of the TFM reconstruction theory and the steps of the proposed de-noising method.

2.1 Phase Space Reconstruction

The TFM learning requires reconstructing a data matrix in the phase space from the observed time series by the Phase Space Reconstruction (PSR) technique. Denote the signal to be analysed by $x(t)$ with N data points. The ith phase point vector in the m-dimensional phase space is given as

$$X_i^m = \left[x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\right] \tag{1}$$

where x_i is the ith data point in the signal $x(t)$, m is the embedding dimension, and τ is the time delay. The embedding dimension is calculated by the Cao's method [8]. In order to keep a high time resolution, the time delay is set to be one [7]. When aligning the vectors $\{X_i^m \mid i = 1, 2, \ldots, n\}$ in the order of time, a time-dependent data matrix $P \in R^{m\times n}$ ($\tau = 1$, $n = N-m+1$) constructed in the phase space is shown as below:

$$P = \begin{bmatrix} X_1^m \\ X_2^m \\ \cdots \\ X_i^m \\ \cdots \\ X_n^m \end{bmatrix}^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ x_2 & x_3 & \cdots & x_{m+1} \\ \cdots & \cdots & \cdots & \cdots \\ x_i & x_{i+1} & \cdots & x_{i+(m-1)} \\ \cdots & \cdots & \cdots & \cdots \\ x_n & x_{n+1} & \cdots & x_N \end{bmatrix}^T \tag{2}$$

In fact, the elements of P and the data points of $x(t)$ has the following relationship:

$$P_{(j,k)} = x_{k+(j-1)\tau} \tag{3}$$

where $j \in [1, m], k \in [1, N - (m-1)\tau]$.

2.2 TFM Analysis and Synthesis

The following problem is calculation and synthesis of the TFM in phase space. Firstly, each row (with the time sense) of the data matrix P is analysed by the STFT to provide a time-frequency representation for the constructed data P in the phase space. The STFT is formulated as bellow:

$$S_j(k, v) = \sum_{l=-\infty}^{\infty} P_j[l] w[k-l] e^{-i\frac{2\pi}{M}vl},\ j = 1, 2 \ldots, m \tag{4}$$

where k and v are the location of time axis and frequency axis, respectively, M is the discrete frequency points number in STFT, $w(k)$ is a short-time analysis window, and P_j is the jth row of matrix P with length n. The result $S(k, v)$ can be written in two parts: amplitude and phase. The amplitude part is generally called time-frequency distribution (TFD). Therefore, we can generate m TFDs from the constructed data P. To improve computation efficiency of the TFM, these m TFDs will receive a process of frequency band selection. That is to say, only the frequency band of interest (e.g., the fault characteristic band) will be kept to form an updated TFD for each one. The m updated TFDs are denoted by $TFD_x^m(t, v)$ with the size of $L \times n$, where L is the selected frequency points, n is the time points. The updated TFDs will be the input of manifold learning.

The manifold learning algorithm, Local Tangent Space Alignment (LTSA) [9], is then employed to calculate the d TFMs, which are conveniently denoted by $TFM_x^d(t, v)$ also with the size of $L \times n$. For the details of the TFM learning, please refer to [7]. Generally, d is far less than m. We can usually take the first TFM as the final TFM result. In order to achieve a better effect, a simple zero threshold processing is employed to the TFM. The TFM has the merit that it can keep the intrinsic time-frequency structure, while the random noise can be restrained.

For the TFM synthesis, the TFM result is used to replace each of m selected TFDs while keeping the relative amplitude. At the same time, the part of TFD without selection for manifold learning is set to be zero. In this way, we can re-generate m TFDs in whole frequency band. Combining the re-generated m TFDs with the m original phase part of $S_j(k, v), j = 1, 2, \ldots, m$, then m updated STFT results, denoted by $\hat{S}_j(k, v), j = 1, 2, \ldots, m$, can be generated.

Then time-frequency synthesis is employed to calculate a new data matrix $\hat{P}$ in phase space by the m updated STFT results. As each STFT result can generate a time series by time-frequency synthesis, m time series could thus construct a data

matrix with the same size of original data matrix P. The time-frequency synthesis of STFT can be expressed as follow:

$$\hat{P}_j[k] = \frac{1}{Mw[0]} \sum_{v=0}^{M-1} \hat{S}_j(k, v) e^{i\frac{2\pi}{M}kv},\ j = 1, 2, \ldots, m \tag{5}$$

By organizing all of the obtained time series, a updated data matrix $\hat{P}$ with the size of $m \times n$ in phase space is formulated.

2.3 PSR Synthesis

After obtaining the updated data matrix $\hat{P}$ in phase space, the PSR synthesis is employed to reconstruct the signal which has de-noising effect. During the process of reconstruction, we should consider the situation that every element of the original time series may appear at several places in the original data matrix in the phase space. The time series reconstructed from the updated data matrix in phase space by PSR synthesis is given as:

$$\hat{x}_i = \frac{\sum_{q \in \{I_i(j,k)\}} \hat{P}_q}{C_i},\ i = 1, 2 \ldots, N;\ j = 1, 2, \ldots, m \tag{6}$$

where $\{I_i(j,k)\}$ is the subscript set of all the elements of the original time series which meet the requirement of $k + (j-1)\tau = i, k \in [1, N - (m-1)\tau]$, and C_i is the number of elements in $\{I_i(j,k)\}$. The final result of the de-noised signal can be thus represented as $\hat{x}(t)$ with N data points.

2.4 Summary of the Proposed de-Noising Method

In summary, the procedure of the proposed de-noising method can be described briefly as follows:

1. Given a signal $x(t)$ with N data points, calculate the data matrix P of size $m \times n$ ($n = N - m + 1$) by PSR.
2. Calculate the STFT of each row of matrix P, $S_j(k, v)$, $j = 1, 2, \ldots, m$, and get its amplitude and phase parts.
3. Select the frequency band of interest to get m TFDs, $TFD_x^m(t, v)$, with the size of $L \times n$.
4. Calculate the TFM of size $L \times n$ by manifold learning, and conduct a zero threshold processing for the TFM.

5. Update the STFT result to get $\hat{S}_j(k,v)$, $j = 1,2,\ldots,m$, using original phase and the TFM as new amplitude.
6. A new data matrix $\hat{P}$ of size $m \times n$ in phase space is calculated by time-frequency synthesis.
7. The de-noised signal $\hat{x}(t)$ is reconstructed by PSR synthesis.

3 Application to Bearing Fault Signal Reconstruction

To confirm the effectiveness of the proposed method, the bearing data with rolling-element defect and outer-race defect are analyzed. Two traditional de-noising methods are also employed for a comparison. They are band-pass filtering method and discrete wavelet transform-based de-noising method. The set of bearing data are from the Case Western Reserve University (CWRU) Bearing Data Center Website [10]. They were acquired by using an experimental setup as shown in Fig. 1. The test stand consists of a 2*hp* motor (left), a torque transducer/encoder (center), a dynamometer (right), and control electronics (not shown). The test bearings support the motor shaft. Vibration data were collected using accelerometers, which were attached to the housing with magnetic bases, with the sampling frequency of 12 kHz for driving end bearing experiments. The bearings used in this test are the deep groove ball bearings with the type of 6205-2RS JEM SKF. Single point defects were set on the test bearings separately at the rolling element and outer raceway using electro-discharge machining. Accelerometers were placed at the 12 o'clock position when the defects were at the rolling element, and at the 6 o'clock position for the outer raceway defect.

3.1 Application to Rolling-Element Defective Signal

The signal with rolling-element defect is firstly analyzed. Figure 2 shows the waveform, power spectrum and the TFD of the defective signal. It can be seen that the waveform indicates a series of similar periodic impulses with the time, but these

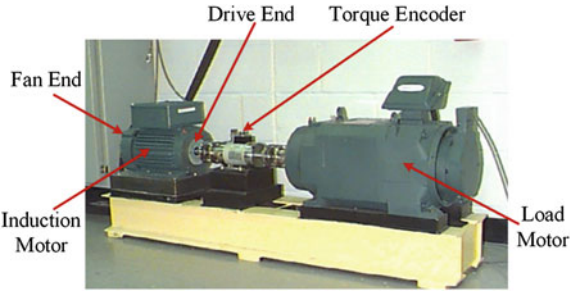


Fig. 1 The bearing test stand

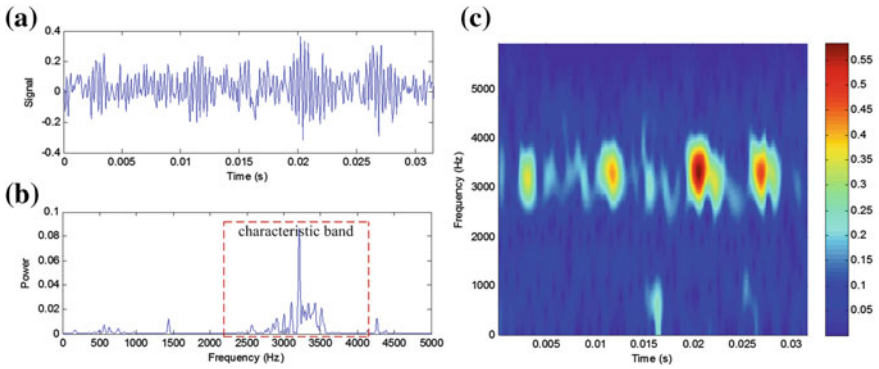


Fig. 2 The rolling-element defective signal: **a** waveform; **b** power spectrum; **c** TFD

periodic impulses cannot be identified directly from the waveform. In the spectrum, the fault characteristic band is roughly estimated around 3,200 Hz. Then we make the frequency band of interest from 2,200 to 4,200 Hz. As shown in Fig. 2c, the TFD presents a combination of time and frequency information. It can be seen that there are a series of impulses along the line of 3,200 Hz. However, there exists the issue of noise corruption.

Firstly, the data matrix in the *m*-dimensional phase space is constructed from the original signal by the technique of PSR. Then the TFD of each row of the data matrix is calculated by STFT. To improve computation efficiency, the parts of the obtained *m* TFDs from 2,200 to 4,200 Hz are remained. Figure 3a displays the TFD of the first row of time series in the selected frequency range. Then an organized 2-D matrix is constructed by all these selected TFDs, and the TFM is obtained by inputting this data matrix to LTSA algorithm. To further improve the effect of the TFM, zero threshold processing is employed. The TFM after threshold processing is displayed as Fig. 3b. By comparing Fig. 3a, b, it can be seen that the situation of noise corruption in Fig. 3b is greatly improved referred to the situation in Fig. 3a. By constructing *m* new amplitude matrixes according to the TFM, and combining them with the *m* original phase matrixes, *m* new STFT results are obtained. After processing the *m* new STFT results with time-frequency synthesis, a new data matrix in *m*-dimensional phase space is generated. At last, a time series is constructed from the new data matrix in phase space by PSR synthesis. The final result as displayed in Fig. 4b presents a series of periodic impulses with the time without noise nearly. The average time period of these impulses is around 0.008 s, which nearly equal to the theoretical value of 0.0071 s. The result confirms that the proposed de-noising method can reduce noise effectively and keep the natural structure of fault signals.

The results of band-pass filtering method and discrete wavelet transform-based de-noising method are displayed as Fig. 4c, d, respectively. Figure 4a is the waveform of original signal. The result in Fig. 4c nearly has no noise reduction effect as compared to original signal. Although the result in Fig. 4d has certain de-noising

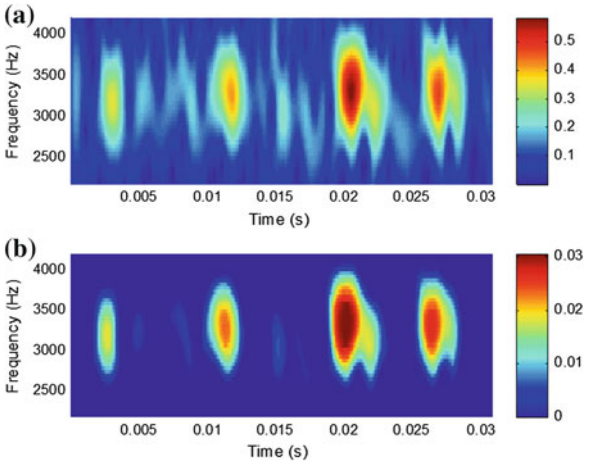


Fig. 3 **a** The TFD of the first dimensional time series in the frequency range; **b** the TFM after threshold processing

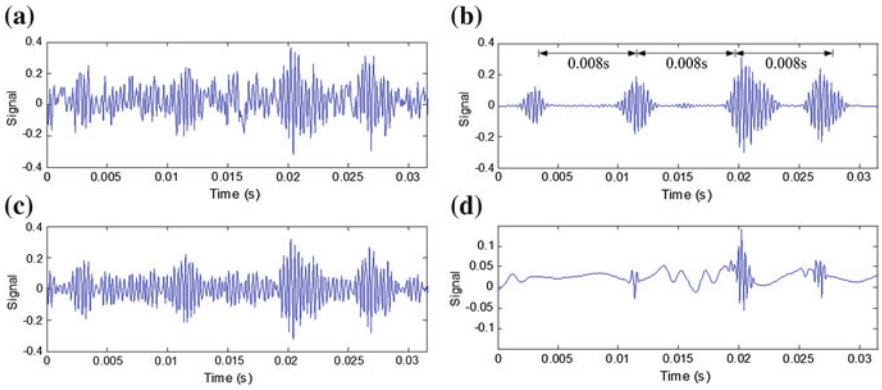


Fig. 4 **a** The waveform of original signal; **b** the result of the proposed de-noising method; **c** the result of band-pass filtering method; **d** the result of discrete wavelet transform-based de-noising method

effect, it does not keep the natural impulse characteristic of original signal. By comparing the four pictures in Fig. 4, it can be found the de-noising effect of the proposed method is obviously superior to the other two methods.

3.2 Application to Outer-Race Defective Signal

Figure 5 shows the waveform, the power spectrum and the TFD of the outer-race defective signal to be analyzed as follows. It can be seen that the existence of noise

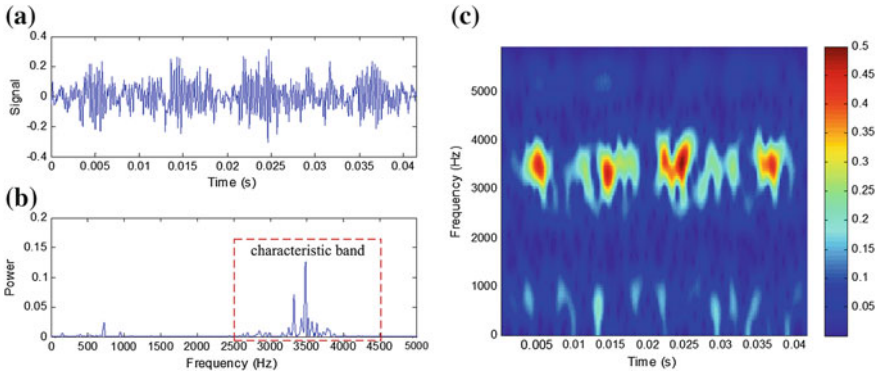


Fig. 5 The outer-race defective signal: **a** waveform; **b** power spectrum; **c** TFD

in the waveform makes it hard to confirm the signal nature. According to the characteristic band in the spectrum displayed in Fig. 5b, the frequency band is determined as [2500, 4500] Hz. Figure 5c shows the TFD, where we can see a series of impulses along the frequency line of 3,500 Hz, but the existence of strong background noise make it hard to confirm the period of these impulses.

After the data matrix in the *m*-dimensional phase space is obtained, the TFD of each row of the data matrix is calculated by STFT. Figure 6a shows the TFD of the first row of time series in the selected frequency range. The obtained TFM with zero threshold processing is displayed in Fig. 6b. It can be seen that the impulses in Fig. 6b are much clearer than those in Fig. 6a. By conducting the time-frequency synthesis, a new data matrix in *m*-dimensional phase space is obtained. The final de-noised signal constructed from the new data matrix by PSR synthesis is presented as Fig. 7b. It can be seen that there are a series of clear periodic impulses. The average time period of these impulses is around 0.01 s, very close to the

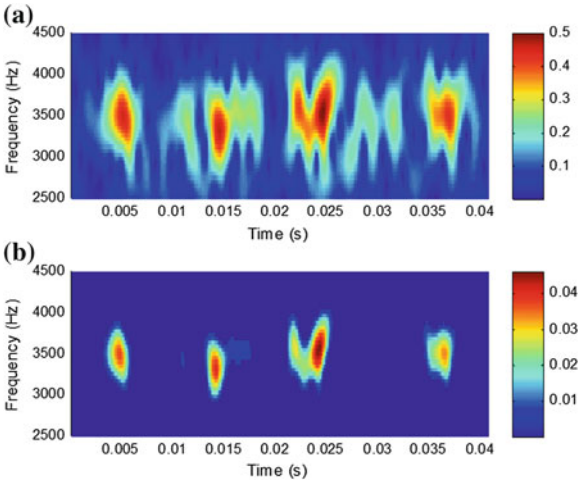


Fig. 6 **a** The TFD of the first dimensional time series in the frequency range; **b** the TFM after threshold processing

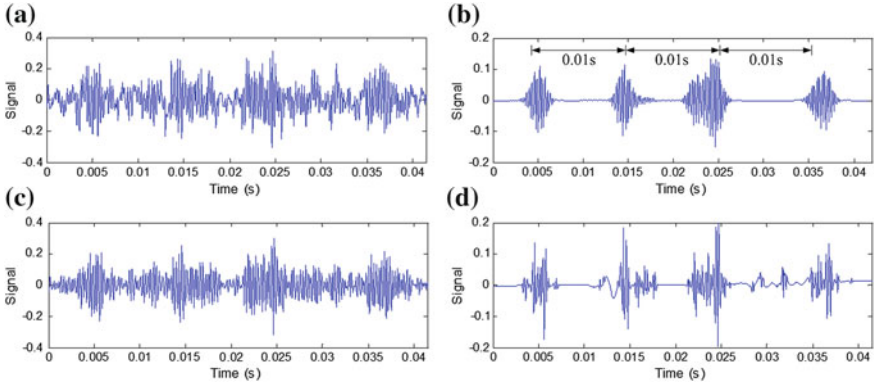


Fig. 7 **a** The waveform of original signal; **b** the result of the proposed de-noising method; **c** the result of band-pass filtering method; **d** the result of discrete wavelet transform-based de-noising method

calculated theoretical value of 0.0096 s. Therefore, the proposed de-noising method is also verified to be able to reduce noise effectively, as well as keep the nature of fault signals.

Two traditional de-noising methods, band-pass filtering method and discrete wavelet transform-based de-noising method, are also used to process this signal. The results are displayed in Fig. 7c, d, respectively. In order to compare conveniently, Fig. 7a, b show the original signal and the result of the proposed method, respectively. By comparing these four pictures in Fig. 7, the de-noising effect of the proposed method is proved to be much more effective than the traditional methods.

4 Conclusion

This chapter presents a novel de-noising method which employs TFM to reconstruct fault signal from the noisy raw signal by combining with the technique of time-frequency synthesis and PSR synthesis. The TFM is introduced to reduce noise for the first time in this study. Thanks to the merits of noise suppression and resolution enhancement of the TFM to display an intrinsic time-frequency structure, the proposed de-noising method can not only reduce background noise effectively, but also keep the intrinsic time-frequency structure of the machinery fault signal, which is significant for a reliable fault diagnosis. The performance of the proposed method has been verified by processing the bearing data with outer-race defect and rolling-element defect in comparison with two traditional de-noising methods including the band-pass filtering method and the discrete wavelet transform-based de-noising method. The results show that the proposed method is rather superior to the two traditional methods in machinery fault signal de-noising.

Acknowledgments This work was supported by the National Natural Science Foundation of China (Grant No. 51005221).

References

1. Papandreou-Suppappola A (2003) Applications in time-frequency signal processing. CRC Press, Boca Raton
2. Deng N, Jiang C (2012) Selection of optimal wavelet basis for signal denoising. In: 9th international conference on fuzzy systems and knowledge discovery (FSKD), pp 1939–1943
3. Beheshti S, Dahleh MA (2005) A new information-theoretic approach to signal denoising and best basis selection. IEEE Trans Signal Process 53:3613–3624
4. Liu L (2011) Using stationary wavelet transformation for signal denoising. In: 2011 eighth international conference on fuzzy systems and knowledge discovery (FSKD), pp 2203–2207
5. Dong X, Yue Y, Qin X, Wang X, Tao Z (2010) Signal denoising based on improved wavelet packet thresholding function. In: 2010 international conference on computer, mechatronics, control and electronic engineering (CMCE), vol 6, pp 382–385
6. Quatieri TF (2004) Discrete-time speech signal processing: principles and practice. China Machine Press, Beijing
7. He Q, Liu Y, Long Q, Wang J (2012) Time-frequency manifold as a signature for machine health diagnosis. IEEE Trans Instrum Meas 61(5):1218–1230
8. Cao L (1997) Practical method for determining the minimum embedding dimension of a scalar time series. Physica D 110:43–50
9. Zhang Z, Zha H (2005) Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. SIAM J Sci Comput 26:313–338
10. http://www.eecs.case.edu/laboratory/bearing/

A Bearing Fault Detection Method Base on Compressed Sensing

Zhang Xinpeng, Hu Niaoqing and Cheng Zhe

Abstract For bearing fault detection in frequency domain, traditional methods estimate bearing fault condition based on mass data sampled by Nyquist sampling theorem, which will burden the storage. A new bearing fault detection method based on compressed sensing will be proposed in this paper in allusion to the problem mentioned above. The method presented here carried out compressive sampling and get a small set of incoherent projections, often the number of projections can be much smaller than the number of Nyquist rate samples. Then based on matching pursuit, the bearing condition will be estimated finally using these few measurements directly without ever reconstructing the signals involved. Sparsity of original signal is not demanded since the signal does not need to be recovered completely, which will also helped to expanded the method to other signals with similar characteristics in frequency domain. Related test will be achieved to verify the effectiveness of the method proposed in this paper.

Keywords Bearing fault detection · Nyquist sampling theorem · Compressed sampling · Matching pursuit

1 Introduction

The bearing is one of the most commonly used but also the most vulnerable part of mechanical equipment, whose working conditions are extremely abominable in addition to high speed and heavy load, and which is a failure-prone component.

Z. Xinpeng (✉) · H. Niaoqing (✉) · C. Zhe (✉)
Laboratory of Science and Technology on Integrated Logistics Support,
College of Mechatronics Engineering and Automation, National University
of Defense Technology, Changsha, China
e-mail: zhangxinpeng@nudt.edu.cn

H. Niaoqing
e-mail: hnq@nudt.edu.cn

C. Zhe
e-mail: chengzhe@nudt.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_69

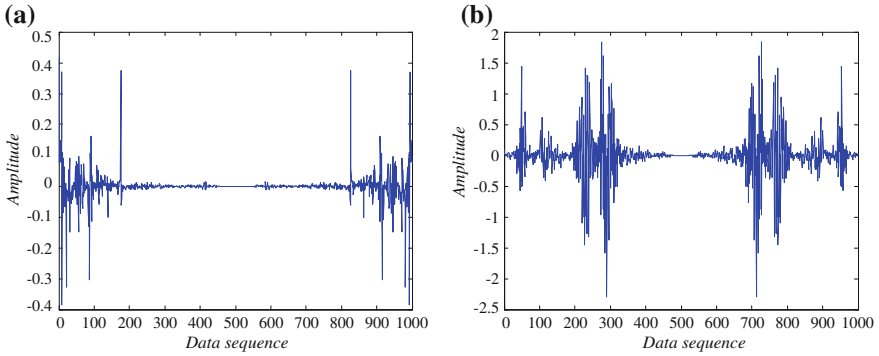


Fig. 1 Typical data sequences after DFT of different bearing states (**a**) normal state (**b**) outer ring fault state

Once failure, it will threaten the safe operation of the equipment, so it is very necessary to monitor the bearings status and identify the fault in time. The vibration caused by bearing faults is characterized by the existence of the shock pulse in the vibration signal. In the time domain, the signal mean and variance etc., will change owing to the shock and in the frequency domain, the shock will level up the high-frequency components so that the energy distribution of the signal changes. Therefore, we can estimate the bearing status by analyzing the energy distribution in frequency domain.

As to bearing fault detection, a relatively simple strategy is to analyze spectrum of the vibration signal we monitored and realize the energy distribution of signals in the frequency domain, then compare the current energy distribution with the energy distribution of known normal state, if the current signal spectrum deviates significantly from the spectrum of normal state, then we can judge that the current bearing condition is abnormal. Suppose $\theta = F \cdot x$, where x is the original signal in time domain, F is DFT (discrete Fourier transfer) matrix, and θ is the data sequence after DCT. Figure 1 shows the data sequence (because θ is a complex vector, Fig. 1 just shows its real parts) after DFT to normal state signal (a) and that of outer ring fault state signal (b) of bearings (6205-2RS JEK SKF deep groove ball bearing, 12 K sampling frequency, 1000 data points, 1797r/min, 1 HP load, and data source from [1]). As can be seen, the energy distribution of the normal state differs obviously from the energy distribution of outer ring fault state, and the intervals contained the data points with maximum amplitude of the normal state and that of fault state are distinct, for example, for normal state, the point with maximum amplitude falls into the interval [1 200] or [801 1000], while for abnormal state, the point with maximum amplitude falls into the interval [201 800]. Therefore, we can use this feature to achieve the bearing fault detection.

When using the ideology mentioned above to detect bearing fault, we have to sample original signals with Nyquist sampling rate and then get all the data points

after DFT to the original signal, and find the data point corresponding to the maximum amplitude, and then combined with prior knowledge, we can finally estimate the bearing condition. The prior knowledge are obtained by comparing the distributions of the data sequence after DCT to original normal state signal and fault signal on the same test condition, also can be described as a learning or training process. Since we just care about the maximum amplitude of data sequence after DCT, then if we use all the data points, the information contained will be redundant. If the number of original data points is extremely large, then we have to save lots of data to achieve fault detection, which will greatly burden the storage. For this problem, we propose a new fault detection method combined with compressed sensing to achieve fault detection using less data points. This new method does not need to obtain large number of data with Nyquist sampling rate but according to compressed sampling, far less data will be obtained, and then based on matching pursuit algorithm, only one iteration will be enough to find the data point with the maximum amplitude in the data sequence after DCT. It is not necessary to completely reconstruct the original signal using this method, and the fewer signals are acquired only for the purpose of making detection. First, we will introduce compressed sensing and matching pursuit algorithm we will use next, and then build the bearing fault detection model based on compressed sampling. The validity and practicability of this method proposed here will be verified by experiments at the end of this paper.

2 Compressed Sampling

As we known, within the framework of traditional digital signal processing base on Nyquist sampling theorem, if we want to recover the analog signal from the discrete sampling signal without distortion, the sampling rate must be at least twice than the signal bandwidth. Currently, most of the data acquisition systems are designed based on the traditional sampling theorem, the data collected in this manner can adequately represent the original signal, but they have great redundancy. Therefore, these methods often result in the collected data proliferation and sensor waste. So it is deeply significant to study how to achieve sampling below the Nyquist frequency according to some characteristics of the signal to reduce the amount of data required to collect.

In 2006, D Donoho, E Candes and T Tao et al. proposed Compressed Sensing (CS) theory [2–6]: Describing the signal using transform space, collecting minority selective linear observation data directly, which contained all the information of the signal, and transferring the signal sampling into information sampling, recovering the original signal from compressed signal by solving an optimization problem. In this theory, the signal sampling rate is no longer dependent on the signal bandwidth, but on the structure and content of information in the signal, so that when these two conditions are satisfied: (1) compressibility of the signal, (2) irrelevance between the representation system and observation system, it will become possible to

recover the high-resolution signal from low-resolution observations [7]. CS avoids the high-speed sampling, which means that the signal sampling and processing can be executed at a very low rate. This will also reduce the data storage and transmission costs significantly and save the signal processing time and computing cost, which will bring new shocks to signal processing. On the other hand, the idea of this compressed observation also leads a new direction for the high-dimensional data analysis. The basic mathematical model of Compressed Sensing is as follows:

If we expand $x \in R^{N\times 1}$ on some orthogonal basis $\{\psi_i\}_{i=1}^{N}$, where ψ_i is an N-dimension column vector, viz.:

$$x = \sum_{i=1}^{N} \theta_i \psi_i \tag{1}$$

where the coefficient $\theta_i = \langle x, \psi_i \rangle = \psi_i^T x$, i.e. representing signal $\boldsymbol{x}$ using a group of orthogonal basis, which can be transferred into a matrix form as:

$$x = \Psi\theta \tag{2}$$

where $\Psi = [\psi_1, \psi_2, \ldots, \psi_N] \in R^{N\times N}$ is as dictionary matrix with orthogonal basis satisfying $\Psi\Psi^T = \Psi^T\Psi = I$, and expansion coefficient $\theta = [\theta_1, \theta_2, \ldots, \theta_N]^T$.

Suppose that coefficient vector θ is K-sparse in basis Ψ, i.e. there has K non-zero elements in θ and $K \ll N$. Using another measurement matrix Φ ($M \times N$ and $M \ll N$)which is irrelevant to Ψ, where each row of matrix Φ can be regarded as a sensor who multiplies with the signal and acquires parts of information of the signal, we can achieve compressed measurement as,

$$y = \Phi x \tag{3}$$

Then we can acquire linear measurement (or projection) $y \in R^M$ which contains enough information to recover signal x, as Fig. 2 shows,

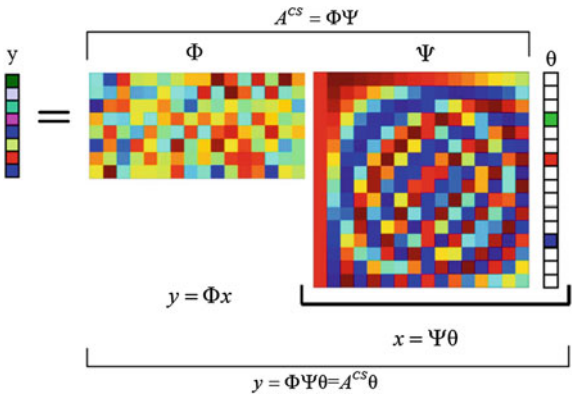


Fig. 2 Matrix representation of compressed sensing

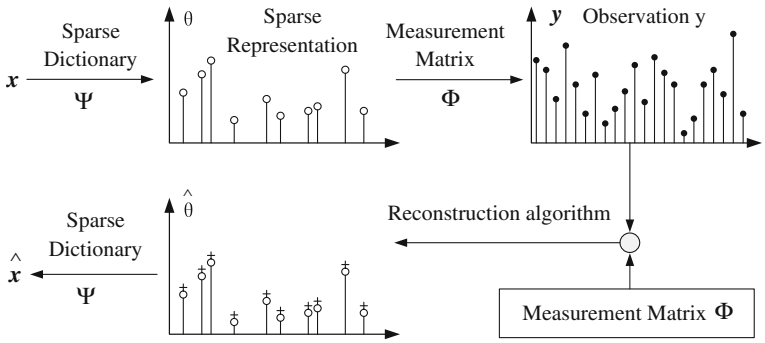


Fig. 3 Process of compressed sensing

There exists a problem of solving linear equations to recover x from y. While Eq. (3) shows that this seems impossible, because this is morbid equation that has more unknowns than equations and has infinite solutions. However, taking Eq. (2) into (3), while we denote $A^{CS} = \Phi\Psi$, then we can get,

$$y = \Phi\Psi\theta = A^{CS}\theta \tag{4}$$

Now, although this is still a morbid problem, while coefficient θ is sparse and then the number of unknowns will greatly decrease, then it can be possible for signal recovery.

Therefore, when signal is or can be represented as sparse and the measurement matrix Φ and the dictionary matrix Ψ is irrelevant, by solving a nonlinear optimization problem, we can reconstruct x from observed y with measurement matrix Φ and dictionary matrix Ψ nearly perfectly. Figure 3 shows the process of Compressed Sensing.

Now we consider the aforementioned problems, for bearing fault detection in frequency domain, the traditional method must collect all the data points in the data sequence after DCT to find the point with maximum amplitude. Based on the principle of Compressed Sensing, we can execute compressed sampling for the original signal to get a small quantity of data points which contained most of information of the original signal. If we can find the location with the maximum amplitude we interested directly using these few data, then the bearing fault detection also can be achieved. In our proposed method, we use DCT matrix as the dictionary matrix Ψ. In fact, we do not need to reconstruct the original signal completely, but only acquire the maximum amplitude information, so we do not demand the original signal is sparse strictly. Then a question emerges that how to get maximum amplitude information from these few data points by compressed sampling. To facilitate the analysis and description later, in the rest of this paper, we denote N as the number of the original data points in time domain, and M as the number of compressed observations ($M << N$). Then the process of compressed

sampling can be expressed as $y = \Phi x$, where $y \in R^M$ is the observed signal and $x \in R^N$ is the original signal, Φ is the measurement matrix. In our method, we carry out DCT as $\theta = F \cdot x$, where F is the DFT matrix, then we can get $y = \Phi x = \Phi \cdot F^{-1} \cdot \theta = A^{CS} \cdot \theta$. So we can describe the problem mentioned above as how to locate the maximum of θ from y directly without recovering original signal x. Matching Pursuit Algorithm provides us a method to solve this problem.

3 Matching Pursuit Algorithm

Matching Pursuit (MP) [8] algorithm is the simplest pursuit algorithm, which is namely as pure greedy algorithm in the estimation theory [9]. The essential idea of this algorithm is selecting the columns of matrix A by greedy iterative approach, viz., every column we selected in each iteration has the highest correlation with current error vector, and then calculating the approximate estimation and the new iterative error using the selected column, repeating the iteration until the number of iterations reaches the threshold we set or the iterative error meet the default error, in which case stop the iteration. Steps of this algorithm are as follows:

Input: observation $\boldsymbol{y}$, measurement matrix **A** and sparsity $\boldsymbol{k}$
Initialization: $\widehat{x}^{[0]} = \mathbf{0}$, $r^{[0]} = y$
FOR $i = 1$; $i = i + 1$; till stopping criterion is met do

Step 1: Calculate the inner product, $g^{[i]} = A^T r^{[i-1]}$;
Step 2: Locate the most important element of $g^{[i]}$, $j^{[i]} = \arg\max_j \left| g_j^{[i]} \right| / \left\| A_j \right\|_2$;
Step 3: Update the estimation $\widehat{x}_{j^{[i]}}^{[i]} = \widehat{x}_{j^{[i]}}^{[i-1]} + g_{j^{[i]}}^{[i]} / \left\| A_{j^{[i]}} \right\|_2^2$;
Step 4: Update the residual error $r^{[i]} = r^{[i-1]} - A_{j^{[i]}} g_{j^{[i]}}^{[i]} / \left\| A_{j^{[i]}} \right\|_2^2$;

END FOR
Output: residual error $r^{[i]}$, estimation $\widehat{x}^{[i]}$

From geometric perspective, MP algorithm first calculates the projections of current residual r to each atom (viz., each column of A), and then find the atom that has the maximum projection, which denoted as j-th column, then this j-th atom is the closest matching atom with the current residual r. Suppose the norm of each atom equals to one, viz., $\|A_i\|_2 = 1$ (i = 1, 2, …, n), then the projection of current residual r to j-th atom is equal to the absolute value of j-th element of x. That is, the following conclusion exists:

If

$$\left|A_j^T \cdot r\right| / \left\|A_j\right\|_2 = \max_{i=1,2,\cdots,n} \left\{\left|A_i^T \cdot r\right| / \left\|A_i\right\|_2\right\}, \quad j \in \{1, 2, \ldots, n\} \tag{5}$$

Then

$$\left|x_j\right| = \max_{i=1,2,\cdots,n} \{|x_i|\} \tag{6}$$

This conclusion is easy to prove, because of $\|A_i\|_2 = 1$, substituting into Eq. (5) we obtain

$$\left|A_j^T \cdot r\right| = \max_{i=1,2,\ldots,n} \left\{\left|A_i^T \cdot r\right|\right\}, \quad j \in \{1, 2, \ldots, n\} \tag{7}$$

by the definition of x, we obtain $x_i = \frac{A_i^T \cdot r}{\|A_i\|_2^2}$, and also because of $\|A_i\|_2 = 1$, then we obtain $x_i = A_i^T \cdot r$, substituting into Eq. (7) we can obtain (6), then conclusion proved.

For our proposed method, here we should set $A = \Phi \cdot F^{-1}$ and θ as the x above. Therefore, the position we located at the first iteration is the same as the location corresponding to element of θ which has the maximum absolute value. Applying MP algorithm, we select the right matrix A, set sparsity of θ as $k = 1$, then the element with largest absolute value can be located after only one iteration, and for other locations, as we set $k = 1$, the algorithm will set all the element of θ as zeros except the location with largest absolute value. Although the original signal x can't be reconstruct in this case, what we are intended to is not reconstructing the original x, but only to find the location corresponding to the largest element in θ. So now a method detecting the bearing fault from compressed observation y directly can be summarized as setting the sparsity of θ as $k = 1$, and then applying the MP algorithm with one iteration to locate the maximum absolute value of θ, then determining whether the bearing has fault according to the information of the location with maximum absolute value of θ combined with prior knowledge. We will build bearing fault detection model based on MP algorithm, and then verify the model with experimental data in the rest of this paper.

4 Experimental Tests

Based on the foregoing analysis, we can build bearing fault detection model based on Compressed Sensing, which is shown in Fig. 4,

And the procedure of fault detection corresponding to the model above is as follows:

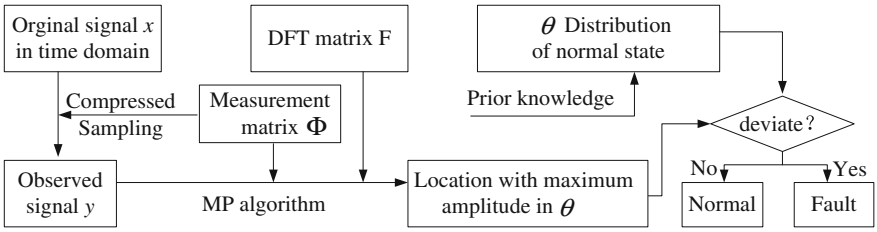


Fig. 4 Bearing fault detection model based on compressed sensing

Step 1: Estimating the θ distribution of normal state based on prior knowledge and set the interval (denote as T) corresponding to normal state;
Step 2: Selecting appropriate measurement Φ;
Step 3: Getting observed signal y based on compressed sampling;
Step 4: Finding the location with maximum amplitude in θ applying MP algorithm based on DFT matrix F and measurement matrix Φ;
Step 5: Detecting the bearing fault by the position of maximum amplitude in θ, viz., if the location corresponding to maximum amplitude in θ falls out of the interval T we set based on prior knowledge, we determine the bearing fault occurs; otherwise, is normal.

We collected vibration data of different states from 6205-2RS JEK SKF deep groove ball bearings in experiments while bearing rev is 1797 rpm and signal sampling frequency is 12 K, 1HP load. Test data contained 960 samples, including 480 normal state samples and 480 fault samples, each of which constituted by 1000 data points. The test data sources are from [1] which is the same as the data we use in part 1. Based on prior knowledge, we set the interval T corresponding to normal state as $T = [1\ 200] \cup [801\ 1000]$. If the bearing condition is normal, then the point with maximum amplitude should fall in the interval T, and if the fault occurs, the point with maximum amplitude will fall out of the interval T. So we can estimate the bearing condition according to the position with maximum amplitude in data sequence θ. In our tests, we used Gaussian random measurement matrix [10, 11], and considering sparsity $k = 1$, according to the principle selecting the number of observational data points [6, 12], we observed 50 points, 100 points and 150 points separately, that is, using measurement matrix Φ of 50×1000, 100×1000, 150×1000 separately.

Now we review the performance of the proposed fault detection method. According to the aforementioned steps, we built the fault detection model based on matching pursuit algorithm to execute detection for the 960 samples. To avoid the instability caused by the randomness of measurement matrix Φ, the tests were repeated 1000 times respectively and used the average value as the final detection result, which is shown in Table 1.

We can see from Table 1 that, with the number of observations increasing, the detection rate increases. The performance of the method we proposed will improve using more observations. That is to say, we just use fewer data points, like 100

Table 1 Fault detection result

Observations	State	Object samples	Detected samples	Detection rate (%)
50	Normal	480	356	74.17
	Fault	480	343	71.46
100	Normal	480	458	95.42
	Fault	480	406	84.58
150	Normal	480	478	99.58
	Fault	480	418	87.08

points to achieve fault detection. In practice, for example, we just need to get and save 100 observations based on compressed sampling to achieve bearing fault detection, while with Nyquist sampling rate we have to get and save 1000 points to achieve the detection on the same test condition. So, for detection problem, the method proposed here can well alleviate sampling pressure facing high frequency signals and storage pressure facing mass data.

5 Conclusions

According to the fact that spectral energy distributions differ between normal bearing signals and fault bearing signals, a bearing fault detection method based on compressed sampling and MP reconstruction algorithm was proposed, and the test results showed that the proposed method can reduce the number of necessary data points significantly while ensuring a well detection rate. In the process of compressed sampling, we used the Gaussian random measurement matrix with different observations. In fact, we can also choose other measurement matrix such as Bernoulli random matrix, etc., so one of the further works will focus on using different measurement matrices and taking different number of observations, in which cases the performances of fault detection algorithms proposed here will be researched. In our proposed method, we only used the information of maximum amplitude, and because the bearings we used in the test has relatively serous fault feature, so we can get a well detection rate. While in the condition without obvious fault feature, the detection result may be instable using only the maximum amplitude information of θ. So in our future work, we will try to use the information of several important amplitudes to achieve detection. Meanwhile, the proposed method is based on the fact that spectral energy distributions differ between different bearings conditions, considering the further application of this method, sparse representation to vibration signal will be also a direction for future work.

Acknowledgments Financial support: This investigation was partly supported by National Natural Science Foundation of China under Grant No. 51075391 and No. 51205401, the Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant No. 20114307110017. Valuable comments on the paper from anonymous reviewers are very much appreciated.

References

1. http://www.eecs.case.edu/laboratory/bearing/fault_specs.htm
2. Donoho D (2006) Compressed sensing. IEEE Trans Inf Theory 52(4):1289–1306
3. Candes E, Wakin M (2008) An introduction to compressive sampling. IEEE Signal Process Mag 25(2):21–30
4. Donoho D, Tsaig Y (2006) Extensions of compressed sensing. Sig Process 86(3):533–548
5. Candes E, Romberg J, Tao T (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theory 52(2):489–509
6. Candes E, Tao T (2006) Near optimal signal recovery from random projections: Universal encoding strategies. IEEE Trans Inf Theory 52(12):5406–5425
7. Licheng J, Shuyuan Y, Fang L, Biao H (2011) Development and prospect of compressive sensing. Acta Electronica Sinica 39(7):1561–1564
8. Mallat S, Zhang Z (1993) Matching pursuit with time-frequency dictionaries. IEEE Trans Signal Process 41(12):3397–3415
9. Friedman JH, Tukey JW (1974) A projection pursuit algorithm for exploratory data analysis. IEEE Trans Comput 23(9):881–890
10. Duarte M, Davenport M, Wak NM et al. (2006) Sparse signal detection from incoherent projection. In: IEEE international conference on acoustics, speech, and signal processing (ICASSP), Toulouse, France, pp 305–308
11. Tropp JA, Gilbert AC (2007) Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans Inf Theor 53(12):4655–4666
12. Baraniuk Richard, Davenport Mark, DeVore Ronald et al (2007) A simple proof of the restricted isometry property for random matrices. Constructive Approximation 23 (4–6):918–925

Implementing IVHM on Legacy Aircraft: Progress Towards Identifying an Optimal Combination of Technologies

Manuel Esperon-Miguez, Ian K. Jennions and Philip John

Abstract The aim of Integrated Vehicle Health Management (IVHM) is to improve the management of maintenance operations through the implementation of health monitoring tools on key components either by diagnosing deterioration or by estimating Remaining Useful Life (RUL) so as to effect timely, and cost effective, maintenance. Regarding the use of IVHM technology in legacy aircraft, one has to keep in mind that hardware modifications to improve the reliability of components is not normally considered a viable alternative to diagnostic and prognostic tools due to high certification costs. At the same time, the data and expertise gathered over years of operating the aircraft help to estimate much more accurately how different health monitoring tools could impact maintenance activities. Consequently, selecting the optimal combination of health monitoring tools for legacy aircraft is significantly easier than for a new design. While computer simulations of the maintenance process are essential to determine how different IVHM tools generate value for the stakeholders, it is not practicable to simulate all possible combinations in order to select which tools are to be installed. This paper describes a process to reduce their number of toolsets to be simulated starting with the identification of those components that present a higher potential to reduce maintenance costs and times in case their faults could be detected and/or predicted. This is followed by the definition of the minimum required accuracy of diagnostic and prognostic tools for each component. This enables designers to determine which tools—available or still being developed—can be implemented to achieve the expected improvement in maintenance operations. Different combinations of IVHM tools are then subjected to a preliminary risk and cost-benefit analysis. A significantly reduced number of combinations are then simulated to select the optimal blend of technologies.

M. Esperon-Miguez (✉) · I.K. Jennions · P. John
IVHM Centre, Cranfield University, Cranfield, Bedfordshire MK43 0FQ, UK
e-mail: m.esperonmiguez@cranfield.ac.uk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_70

1 Introduction

Integrated Vehicle Health Management (IVHM) aims to maximise the use of an asset and reduce its through life maintenance cost through the implementation of health monitoring tools that generate information regarding the condition of multiple components. This information is generated by either diagnostic or prognostic tools. Diagnostic tools reduce the time necessary to detect and isolate a fault, and can be used to avoid human error in the identification of faulty components. Prognostic tools estimate the Remaining Useful Life (RUL) of the component which, at least, helps to avoid a failure during a flight, allowing for the mission to be completed successfully and avoiding any secondary damage. If an accurate prognosis can be generated with enough time in advance (a.k.a. prognostic window or lead time), the replacement of the component can be scheduled at a time and location that minimises—or avoids—any disruption in the operation of the vehicle.

The implementation of IVHM technology has traditionally followed a reactive approach according to which a health monitoring tool is developed individually and, once its performance has been tested, it is put into service. There are two explanations for this approach: on one hand diagnostic and prognostic algorithms and the hardware necessary to implement them are normally developed independently by teams with expertise in the component/system being monitored; on the other hand, organizations lack a high level IVHM policy or program that requires a comprehensive analysis of the optimal combinations of tools to be developed and implemented. Consequently, aircraft end up with an eclectic set of IVHM tools that improve the maintainability of each part, but may have a negligible effect on the availability of the fleet.

However, it must be noted that the lack of a systems approach to IVHM implementation is not caused by lack of competence or vision. The use of several tools on a given aircraft results in interactions that must be carefully studied to ensure objectives are reached and their performance not undermined by overseeing critical interdependencies. From a maintenance perspective it is essential that the selection of components to be monitored takes into account their failure/replacement frequency, replacement time, delays and how IVHM can affect them. Given the complexity of maintenance operations this problem must be studied using computer-based simulations. From an implementation perspective, the interactions between tools can result in unforeseen problems with the hardware and/or the software. Thus, implementing and IVHM system that comprises diagnostic and prognostic tools that monitor several component becomes an engineering project that requires a significant investment and involves a great uncertainty.

Some methodologies to approach this problem do exist, but they normally focus on individual parts or a limited number of components or subsystems. It has been proposed to use Failure Modes, Effects and Criticality Analysis (FMECA) as the main basis for the design of full IVHM systems [1, 3, 9]. However, these methodologies, while applicable to a limited number of components, are not suitable for the analysis of a complete aircraft since it would be impractical to carry out a

FMECA for each individual part, not to mention the analysis of all possible interactions between components and between their potential monitoring tools. As it is explained in the following sections, in the case of legacy aircraft, their unique combination of abundant historical maintenance data and constraints that rule out significant modifications of their systems, allow for a series of quantitative analyses leading to an optimal combination of diagnostic and prognostic tools.

Although IVHM can include the use of tools to improve the management of logistics, maintenance and operations, this paper discusses a methodology to select the optimal combination of diagnostic and prognostic tools by performing different quantitative analyses before defining the final set of tools based on the results obtained from a computer-based simulation of maintenance activities. Consequently, in this text the use of the terms "health monitoring tools" or simply "tools" makes reference to diagnostic or prognostic tools.

2 IVHM and Legacy Systems

Retrofitting IVHM into legacy platforms presents a very specific set of challenges that must be acknowledged from the beginning. While some of these issues affect all kinds of aircraft, they are more acute for aircraft that have been operated for years but are no longer being manufactured. A short list and discussion are presented bellow to show the breadth and depth of these issues.

Technical constraints: Geometric and weight constraints can result in the need to make changes to the structure or other components to accommodate new sensors, wires, electronics, etc. However, the cost of certifying the new tools and any changes required can exceed that of the design, manufacturing and installation of the necessary hardware. The need to ground the aircraft to install and test any new IVHM tool can disrupt normal operations and result in a loss of revenue, making these modifications even more difficult to justify. Software faces similar challenges given the critical role it plays nowadays both on-board and off-board [4]. The cost of certifying major hardware modifications and the uncertainty of potential benefits undermine the implementation of IVHM technology on legacy aircraft [2]. Consequently, for health monitoring tools to be implemented they must require very small or no modifications of existing systems.

Role of organizations: The implementation of IVHM has a significant impact on each stakeholder's organization and vice versa. Aircraft will have to remain grounded for a significant amount of time resulting in significant disruptions in normal operations [4]. Moreover, in order to maximise the benefit of this technology maintenance practices have to change to be able to act based on the information provided by the new health monitoring system. Cultural barriers such as lack of understanding of the real benefits of IVHM and insufficient management support can jeopardize its development and put in service [7, 11].

Regulations and standards: Maintenance organizations normally have some Condition-Based Maintenance (CBM) policies already in place. Depending on the

aircraft, the organization it belongs to and its area of operation some of these procedures can be regulated and made compulsory. As a consequence, a prognostic tool that monitors a component for which CBM is compulsory is not likely to be justifiable from an economic stand point since the investment will not be translated into a significant saving [5]. Therefore, special attention must be paid to maintenance regulations and standards.

Historical Maintenance Data: What sets legacy aircraft apart is the amount of information regarding the reliability of their components, operational environment, maintenance processes and failure modes. While new aircraft rely on estimates based on their design characteristics or a few tests, legacy aircraft present much more comprehensive datasets with information gathered in real operational conditions. As a result, not only is there more information available, but also it is much more accurate.

Although there is a lot of information recorded in maintenance and mission logs it can be difficult to transform it into useful data for the development of an IVHM system. Analysing records kept in handwritten documents or early databases can become an arduous task. Nevertheless, this still represents a significant advantage over new aircraft and, as will be discussed in the following sections, proves crucial in the selection of the optimal combination of health monitoring tools to be retrofitted on a given aircraft.

2.1 Identifying the Role of Stakeholders

Given the complexity of the aviation industry nowadays, the role of different stakeholders must be identified from a very early stage. Whereas in the past the owner, operator and maintainer of a fleet were the same entity, outsourcing and leasing have generated all sort of different sources of revenue, but also makes it difficult to pinpoint who should pay for the development of IVHM technology. Furthermore, health monitoring technology can underpin the transformation of manufacturers to service providers, meaning any CBA for IVHM must take into account the effect it can have on current and future contracts as well as the company's mid and long-term strategy [8]. Wheeler et al. [13] identified the goals for different stakeholders according to their responsibilities: logistics, mission operation, maintenance and fleet management. These goals are then divided into those which can be achieved using diagnostic tools and those which need the use of prognostic tools.

2.2 Framing the Problem

The fact that major modifications of a legacy aircraft's systems are too expensive represents an advantage compared to new aircraft for which this is a viable option. For legacy aircraft, the business case for an IVHM system to monitor a certain

group of components is very easy to justify when faced with the option of modifying such components to improve their reliability and maintainability to a level that results in the same improvement in cost and availability. Consequently, these limitations can be seen as the constraints for a mathematical problem in which major changes in an aircraft's systems are no longer an option. As a result, it can be assumed that the performance of the aircraft is not going to be affected, nor will its interdependencies between systems.

Computer simulation of aircraft maintenance systems can be used to study how health monitoring technology affect maintenance activities and, consequently, maintenance cost and availability at aircraft and fleet level. Unlike aircraft that are being designed or have only been operated for a short period, legacy aircraft can rely on historical maintenance data to provide all the information necessary for the development of these models.

In summary, the use of historical maintenance data in combinations with the constraints just mentioned helps to formulate accurate CBAs for IVHM systems for legacy aircraft.

3 Quantifying the Benefits of IVHM

IVHM affects both maintenance costs and times. Consequently, the availability of an aircraft—and the squadron and fleet to which it belongs—will depend on the tools that form such an IVHM system. Not only does health monitoring reduce the time necessary to replace a component by performing faster diagnoses or avoiding secondary failures, but it can also affect the timing of, and location for, maintenance actions. Taking into account that several tasks are performed simultaneously during both scheduled and unscheduled maintenance stops, it is not possible to calculate analytically the duration of each stop. Furthermore, delays play a major role in maintenance and can be due to different logistic, administrative or technical causes. The fact that maintainers organise maintenance tasks depending on operational demands and the minimum equipment lists for future missions only increases the complexity of the problem. It is only through the use of computer-based simulations of maintenance activities that these complexities can be captured and the effect of IVHM technology estimated quantitatively.

The development and validation of these models requires significant amounts of data. To ensure the benefits of implementing IVHM are estimated correctly these datasets must include, not only the average of variables such as Mean Time Between Failures (MTBF) or Mean Time To Replace (MTTR), but also their variances.

Evidently, the model must take into account the effect any potential diagnostic or prognostic tool can have on maintenance costs and times as well as availability. In order to do so it is essential to acknowledge that health monitoring tools are not 100 % accurate. Diagnostic tools can produce false positives (a.k.a. false alarms) by indicating a healthy component has failed, or false negatives if a faulty component

is not detected. Similarly, prognostic tools estimate the RUL of a component at certain point in time and its replacement is scheduled according to that estimate, but if the estimation is too optimistic it might have to be replaced at a less convenient time and location or even fail during a flight. Being able to simulate the performance of health monitoring tools is essential to compare tools with lower cost and performance with more reliable and expensive ones.

3.1 Reducing the Number of Runs

Ideally, once the maintenance model has been developed and validated, different combinations of diagnostic and prognostic tools can be tested on it. However, while the computer model is the only way to carry out a solid CBA, it is not practical—or even possible—to simulate the effect of all potential combinations. Taking into account that aircraft comprise thousands of components, a comprehensive analysis of all options should consider, at least, a few dozen components to be monitored, even if the final number of tools to be implemented may be lower. For example, if 10 tools are to be chosen out of 50 possible options, this represents more than 10 billion possible combinations. Even taking into account incompatibilities between tools due to conflicts caused by their hardware or software, it is unlikely that the total number of toolsets is reduced significantly enough so all combinations can be studied and compared thoroughly.

Consequently, there is a need for a methodology to reduce the number of combinations of diagnostic and prognostic tools whose impact on maintenance cost and availability is to be studied using a computer simulation of aircraft maintenance activities. Such methodology must be based on a set of quantitative analyses to avoid any bias. Several combinations must be generated with this methodology to allow for sanity checks and to compare how they affect other factors apart from cost and availability. This methodology has been developed taking into account the constraints imposed on legacy aircraft, the availability and accuracy of historical maintenance data and the information that can be gathered at the conceptual design stage on the characteristics and performance of health monitoring tools. The main steps, which will be discussed in detail in the following sections, are:

1. Identify components more likely to have their maintenance time and cost reduced if monitored.
2. Select a preliminary list of health monitoring tools capable of detecting or predicting the failure of the components previously selected.
3. Identify incompatible combinations of tools due to software or hardware conflicts.
4. Preselect toolsets according to their expected Return On Investment (ROI) and financial risk.

3.2 Identifying Critical Components

The first step to reduce the number of simulations necessary for a comprehensive comparison of all the alternatives for an IVHM system involves identifying which components should be monitored. At this phase the number of components preselected is larger than the number of parts that will finally be monitored to allow for modifications in later stages. The objective is to identify which components are more likely to reduce maintenance time and cost if they are monitored by a diagnostic or a prognostic tool.

It is easy to evaluate what is the cost of replacing each component per flying hour as well as its corrective or preventive maintenance time per flying hour. Diagnostic tools essentially reduce the time dedicate to fault identification and isolation which will only affect labour costs. A prognostic tool affect the probability of a component having to be replaced at different locations (affecting logistic delays and shipping costs) and whether it will be an unscheduled task or part of a scheduled maintenance stop (with different costs and delays).

A method proposed in the past consists of analysing the possible outcomes of failure using Event Tree Analysis (ETA) [6] (Fig. 1). Using the probability of a certain component failing as starting point, the tree forks based on the outcome of using a certain type of IVHM tool. A long-term prognostic tool can provide a RUL with a prognostic window long enough to schedule the replacement of the part so it will not affect the inherent availability of the aircraft. However, the estimated RUL can be incorrect. In that case there is the possibility to use a short-term prognostic algorithm and replace the part during an unscheduled stop, avoiding a possible mission loss or even secondary damages. Nevertheless, this recalculated RUL can

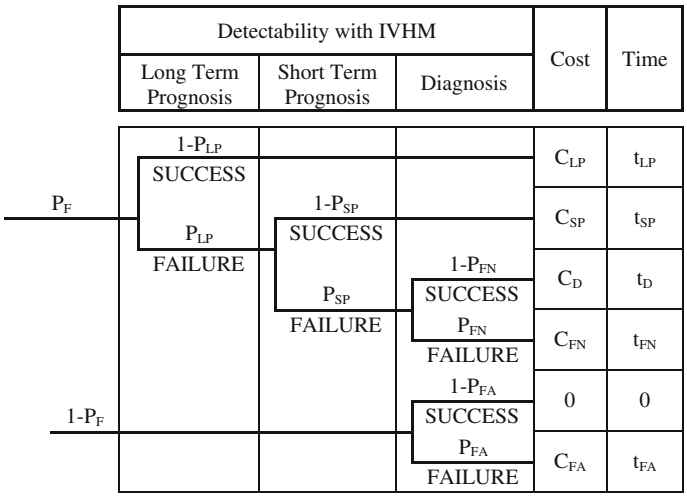


Fig. 1 ETA for the use of health monitoring tools on a single component

also be overly optimistic, meaning the failure will take place and will need to be detected and isolated. If a diagnostic tool is fitted this can be performed automatically, but there is always the possibility of a false negative resulting in a longer time to diagnose the fault. The tree also includes the possibility of a healthy component being flagged as faulty by a diagnostic tool.

The order in which these tools appear in the tree does not reflect how health monitoring algorithms operate, it simply indicates in which order they will define the final outcome. Additionally, the fact that three kinds of tools are included in the tree does not imply that each component counts with all of them.

One of the advantages of this setup is that it accounts for the fact that some components can utilise some diagnostic capability in the form of Built-In Test Equipment (BITE) or be replaced according a preventive maintenance scheme, which, for the purpose of the ETA, has the same effect as a prognostic tool. Since it is possible to upgrade a health monitoring tool, there is no reason to exclude them from this analysis.

Analytical equations for the maintenance time, T, and costs, C, incurred per flying hour for each component are easy obtained based on this ETA [6]. Since they are polynomial expressions the derivatives can also be calculated analytically quite easily.

$$C = P_F((1-P_{LP})C_{LP} + P_{LP}((1-P_{SP})C_{SP} + P_{SP}((1-P_{FN})C_D + P_{FN}C_{FN}))) + (1-P_F)P_{FA}C_{FA} \tag{1}$$

$$T = P_F((1-P_{LP})t_{LP} + P_{LP}((1-P_{SP})t_{SP} + P_{SP}((1-P_{FN})t_D + P_{FN}t_{FN}))) + (1-P_F)P_{FA}t_{FA} \tag{2}$$

where the performance of long and short term prognostic tools is defined by P_{LP} and P_{SP}, respectively; and the probability of false alarms and false negatives by P_{FA} and P_{FN} respectively. If a long term prognostic tool works correctly the cost and time of replacing the components are C_{LP} and t_{LP} respectively, and C_{SP} and t_{SP} in case a short term prognostic tool is used. If the fault is detected by a diagnostic tool the cost and downtime will be C_D and t_D. For false alarms costs and downtimes are denoted by C_{FA} and t_{FA}, and by C_{FN} and t_{FN} for false negatives.

This ranking takes into account the maintenance time spent on individual components. As it has been discussed previously, there is not a direct correlation between the reduction of maintenance time of certain individual parts and the availability of the aircraft. However, components with longer maintenance times and higher sensitivities to the use of IVHM are more likely to have an important role in the improvement of the availability of the fleet. Once the components have been ranked the computer model can be used to verify which of those at the top of the list are responsible for most of the unscheduled maintenance stops and delays.

Identifying which components are the best candidates to be monitored by diagnostic and prognostic tools is useful, but it is not the kind of information that

can be used to run computer simulations. The model uses the performance of health monitoring tools, meaning tools capable of assessing the condition of these top components have to found, as explained in the following section.

3.3 Performance Requirements for a Preliminary Selection of Health Monitoring Tools

Once key components have been identified it is necessary to find which tools can be used to monitor them. Original Equipment Manufacturers (OEMs), companies specialised in health monitoring technology and universities can be contacted to determine which tools are available or can be developed.

Even at such an early stage in the design of an IVHM system it is necessary to define basic technical and economic requirements to be able to compare different toolsets. Once again, the computer model is essential to define the minimum expected reductions in maintenance times and costs for each component to achieve the desired availability and total maintenance cost.

As shown in Eqs. (1) and (2), it is possible to define the maintenance cost, and time of a component as a function of the performance of different health monitoring tools. If cost and time become a design requirement (C^* and T^* respectively) these equations can be used to define the required performance for a diagnostic (Eqs. 3; 4–6) or prognostic tools (Eqs. 7–9).

The progress in health monitoring technology has not been homogeneous for all kind of systems and it is possible that for certain components diagnostic or prognostic tools that satisfy the performance requirements are not available yet. The possibility of developing a new tool, or improving on an existing one, should be studied at this stage. Conversely, it is also possible that for other components several candidates can be identified. Rather than select a single tool for each component by a process of elimination, all possible options should be considered. The following sections illustrates how the interactions between tools can be studied to identify which components should finally be monitored and by which tool.

Diagnostic Tools

$$P_{FA} \geq 0; \;\; P_{FN} \geq 0 \tag{3;4}$$

$$P_{FA} \leq \frac{C^* - P_F((1 - P_{FN})C_D + P_{FN}C_{FN})}{(1 - P_F)C_{FA}} \tag{5}$$

$$P_{FA} \leq \frac{T^* - P_F((1 - P_{FN})t_D + P_{FN}t_{FN})}{(1 - P_F)t_{FA}} \tag{6}$$

Prognostic Tools

$$P_{LP} \geq 0 \tag{7}$$

$$P_{LP} \leq \frac{\frac{C^* - (1-P_F)P_{FA}C_{FA}}{P_F} - C_{LP}}{(1-P_{FN})C_D + P_{FN}C_{FN} - C_{LP}} \tag{8}$$

$$P_{LP} \leq \frac{\frac{T^* - (1-P_F)P_{FA}t_{FA}}{P_F} - t_{LP}}{(1-P_{FN})t_D + P_{FN}t_{FN} - t_{LP}} \tag{9}$$

3.4 Uncertainties and Their Effect on CBAs for IVHM

Most parameters in maintenance activities are normally random variables due to the fact that even repetitive tasks seldom take the same amount of time or require the same amount of attention and resources. It is possible to work with average values for some basic analyses, but if the objective is to ensure availability stays above a certain value and maintenance costs do not exceed a given limit working with average values results in a 50 % chance of failing to reach the objectives.

The sources of uncertainty can be divided into two main categories. Epistemic, or systemic, uncertainties are caused by inaccuracies in the measurement, recording or modelling of a given parameter. These are the kind of uncertainties which affect the accuracy of maintenance records. To begin with, recorded times are never perfectly accurate but rounded to the nearest multiple of five, ten or 15 min. Additionally, while the total maintenance time spent on each component is often recorded, this is not always the case for the different steps involved (e.g.: preparation, diagnosis, check-out, etc.) or the delays. Even in those few cases when records include this information values are most likely approximations written down after the work has been completed.

The second group comprises the sources of aleatoric, or statistical, uncertainties which are those caused by the random variation of parameters over time. Recurring costs, time spent on different activities, delays and the performance of health monitoring tools are the most prominent. While the amount a supplier charges for a part can be fairly constant (this does not apply to expensive components with low failure rates and low stock), shipping and storage costs can vary considerably. The same can be said about the time dedicated to maintenance tasks, whose variability is related to the complexity of the task.

The uncertainty of the performance of IVHM tools has been well documented. Lopez & Sarigul-Klijn [10], showed how the reliability of an IVHM tool varies depending on the characteristics of the fault, which are different on every occasion, and this translates into uncertainty about its performance. Furthermore, Saxena et al. [12] also analysed how the accuracy of prognostic algorithms evolves with time, with the RUL becoming more accurate as the component approaches its point of failure.

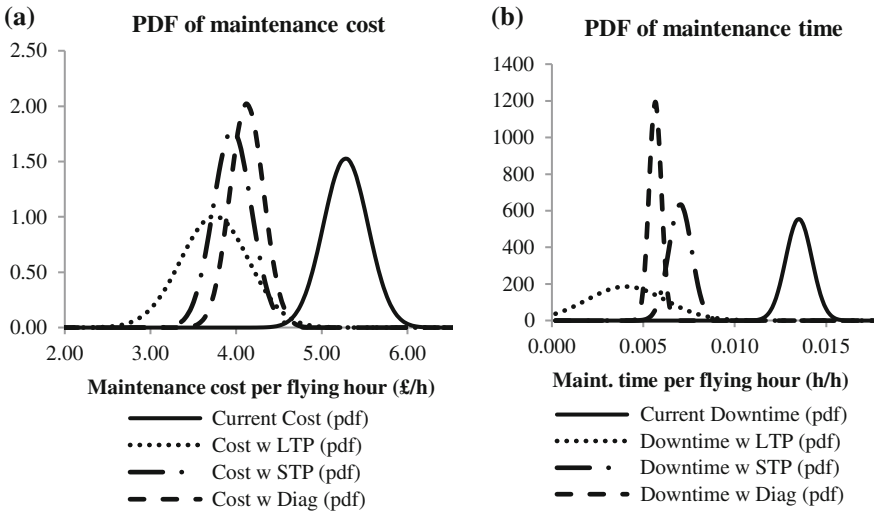


Fig. 2 Examples of the effect of IVHM tools on the PDF of maintenance (**a**) cost and (**b**) time of a component

As a result, engineers who define the performance requirements not only must acknowledge that expected maintenance costs and times follow probability distributions but also take into account that the variance of the performance of each tool must be below a certain threshold. This threshold can be defined using Eqs. (1) and (2) as a basis to determine the variances of performance parameters:

$$Var(C) = Var(P_{FN}P_F(C_{FN} - C_D)) + Var(P_F C_D) + Var(P_{FA}(1 - P_F)C_{FA}) \quad (10)$$

$$Var(T) = Var(P_{FN}P_F(t_{FN} - t_D)) + Var(P_F t_D) + Var(P_{FA}(1 - P_F)t_{FA}) \quad (11)$$

Figure 2 shows the effect diagnostic tools, short term prognostic tools and long-term prognostic tools can have in the probability distributions of the maintenance cost and time per flying hour of a component.

It would seem as if these uncertainties add complexity to our problem increasing the difficulty of finding an optimal combination of diagnostic and prognostic tools. However, as explained in the following section, these uncertainties can be used to carry out a risk analysis of the different sets of tools and to reduce the number of combinations that should finally be studied using computer simulation.

3.5 Balancing ROI and Risk

Comparing toolsets must take into account the possibility of sharing resources between tools in their design, testing, manufacturing, implementation and operation. In other words, tools can share -among others- sensors, memory, flight test

expenses, recurring costs, etc. This translates to a reduction in the investment necessary to put a certain group of tools in service. Consequently, the ROI of each toolset is not the weighted average of the ROIs of those tools it comprises, but the ratio between the sum of their expected profits and the total cost of developing, implementing and operating the complete IVHM system. This profit is essentially based on the costs avoided thanks to the use of a certain health monitoring tool, but other benefits can be included. In mathematical terms, for a toolset with n tools in which the project budget for each tool has been divided into m phases or parts this can expressed as:

$$ROI = \frac{\sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} C_i} = \frac{\sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} \sum_{j}^{m} c_{ij}/\alpha_{ij}} \tag{12}$$

where P_i is the expected profit from tool i; C_i the total cost of tool i; c_{ij} the cost of tool i for part j of its budget; and α_{ij} the number of tools with which c_{ij} is shared.

However, sharing resources means that a deviation in their cost can effectively raise the cost of several tools. For example, if algorithms are processed in a centralised unit whose costs exceeds the original budget this will also impact the cost of each individual health monitoring tool. A federated IVHM system with algorithms run in individual processing units may be more expensive, but its total cost is less vulnerable to this kind of problems.

Comparing toolsets becomes even more complicated when options include tools that are under development and not fully proven. Mature diagnostic and prognostic tools are less likely to present problems and have significant cost variations, but their performance can be lower than tools that are still being developed and employ the latest technology. The cost of the latter, however, is more likely to deviate from the budget.

This resembles a classic financial investment problem in which investors must select the optimal combination of assets to maximise the return of their portfolio while keeping risk within reasonable limits. As in the problem described in this article, financial assets have some degree of correlation and this must be carefully studied to avoid situations in which an investor can be severely affected by fluctuations in the market (e.g.: stock prices of logistic companies are affected by the fluctuation of oil prices in commodity markets, gold prices and the U$D are normally inversely correlated, etc.).

There are all sorts of financial analysis tools that can be applied to solve this problem, but there is an important part of this financial analysis tools ignore: the variation of the ROI of each health monitoring tool depending on how it is combined with others. This is due to the fact that the return on a financial product is not affected by how much one invests in other assets.

Figure 3a shows the result of using a tool known as the efficient portfolio frontier to analyse combinations of IVHM tools. As toolsets include larger numbers of diagnostic and prognostic tools the risk decreases because deviations in the cost of individual tools have a smaller impact on the total investment. However, the ROI

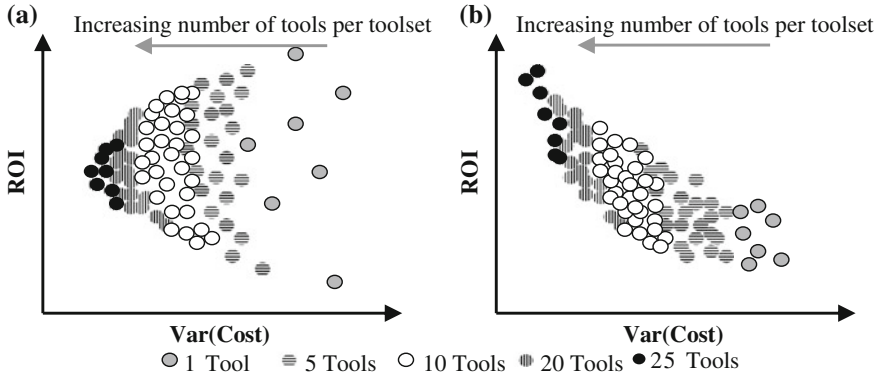


Fig. 3 ROI versus variance of cost for IVHM toolsets using financial analysis (**a**) and including shared costs (**b**)

tends to the average ROI of all possible options because the savings are not taken into account. Figure 3b shows how the ROI can increase significantly if IVHM tools are combined appropriately taking into account Eq. (12).

Those toolsets that present a higher ROI and a lower variance of costs can be tested on the computer simulation. This will determine which combination of tools should be retrofitted on the aircraft and provide a much more accurate estimation of the final outcome.

4 Conclusions

The methodology presented in this paper illustrates how it is possible to carry out exhaustive quantitative analyses of the effect of retrofitting different IVHM toolsets on legacy aircraft without being overwhelmed by the number of options to compare. While computer simulations are essential to ensure CBAs for IVHM are accurate, they cannot be the only tool available to define the optimal combination of health monitoring tools.

Uncertainty plays a major role in the analysis and comparison of different toolsets. Design teams must be aware of the main sources or uncertainty and to what degree it affects the information generated at each stage of the process. Second order uncertainties or "uncertainty of uncertainties" is a major area of research IVHM developers cannot ignore. The trustworthiness of any CBA is directly affected by the variance of the variables it uses and to be able to define them a deep knowledge of aircraft design, maintenance and operations is required.

While financial analysis tools can be used to determine how risk changes depending on how diagnostic and prognostic tools are combined, they must be modified to take into account the effect potential savings have on the resulting ROI.

Acknowledgements This work has been supported by the IVHM Centre at Cranfield University. The authors would like to thank the partners of the IVHM Centre for their support in this project.

References

1. Ashby MJ, Byer RJ (2002) An approach for conducting a cost benefit analysis of aircraft engine prognostics and health management functions. In: Proceedings of the 2002 IEEE aerospace conference, vol 6, pp 6–2847
2. Azzam H, Beaven F, Gill L, Wallace M (2004) A route for qualifying/certifying an affordable structural prognostic health management (SPHM) system. In: Proceedings of the 2004 IEEE aerospace conference, vol 6, pp 3791
3. Banks J, Reichard K, Crow E, Nickell K (2009) How engineers can conduct cost-benefit analysis for PHM systems. Aerosp Electron Syst Mag IEEE 24(3):22–30
4. Dunsdon J, Harrington M (2008) The Application of Open System Architecture for Condition Based Maintenance to Complete IVHM, In: 2008 IEEE aerospace conference, pp 1
5. Esperon-Miguez M, John P, Jennions IK (2012) The effect of current military maintenance practices and regulations on the implementation of Integrated Vehicle Health Management technology. A-MEST' 12, Seville
6. Esperon-Miguez M, John P, Jennions IK (2012) Uncertainty of performance metrics for IVHM tools according to business targets. In: First european conference of the prognostics and health management society 2012, vol 11, Dresden, pp 11
7. Grubic T, Redding L, Baines T, Julien D (2011) The adoption and use of diagnostic and prognostic technology within UK-based manufacturers. Proc Inst Mech Eng Part B J Eng Manuf 225(8):1457–1470
8. Hess A, Calvello G, Frith P, Engel SJ, Hoitsma D (2006) Challenges, issues, and lessons learned chasing the " Big P": real predictive prognostics part 2. In: IEEE 2006 Aerospace Conference, pp 1
9. Kacprzynski GJ, Roemer MJ, Hess AJ (2002) Health management system design: development, simulation and cost/benefit optimization. In: Proceedings of the 2002 IEEE Aerospace Conference, vol 6, pp 6–3065
10. Lopez I, Sarigul-Klijn N (2010) A review of uncertainty in flight vehicle structural damage monitoring, diagnosis and control: Challenges and opportunities. Prog Aerosp Sci 46 (7):247–273
11. MacConnell JH (2007) ISHM & Design: A review of the benefits of the ideal ISHM system. In: IEEE 2007 Aerospace Conference, pp 1
12. Saxena A, Celaya J, Saha B, Saha S, Goebel K (2010) Metrics for offline evaluation of prognostic performance. Int J Prognostics Health Manage 1(1), 20, 73
13. Wheeler K, Kurtoglu T, Poll S (2009) A survey of health management user objectives related to diagnostic and prognostic metrics

A New Method of Acoustic Signals Separation for Wayside Fault Diagnosis of Train Bearings

Ao Zhang, Fang Liu, Changqing Shen and Fanrang Kong

Abstract For the acoustic signal acquired by a microphone is composed of a number of train bearing signals and noises, single signal of failure train bearing should be extracted to diagnose the fault type precisely in wayside fault diagnosis of train bearings. However, the phenomenon of Doppler distortion effect in the acoustic signal acquired with a microphone leads to the difficulty for signal separation. In this chapter, a new method based on Dopplerlet transform, time-frequency filtering and inverse generalized S-transform is proposed to separate different fault types of train bearing signals from the acoustic signal. Firstly, search the parameters space to find the primary functions-Dopplerlet atoms which match the original signal best by matching-pursuits-based Dopplerlet transform. According to the parameters, these Dopplerlet atoms are divided into different groups corresponding to diverse acoustic sources. Through extracting the data of Dopplerlet atoms in a group and its neighborhood in time-frequency domain, the signal of corresponding train bearing can be reconstructed by the inverse transformation of GST. To diagnose the fault type of the reconstructed signal, re-sampling is carried out to remove the Doppler distortion effect in advance. After that, we can identify the fault type of reconstructed signal corresponding to a certain train bearing through the envelope spectrum. Finally, experiments with practical acoustic signals of train bearings with a defect on the outer race and the inner race are carried out, and the results verified the effectiveness of this method.

1 Introduction

The defect of the roller bearing is the dominant type of fault for a train, which lead to serious accidents and significant costs for the rail transport industry [1]. The wayside Acoustic Defective Bearing Detector (ADBD) system [1] was developed in

A. Zhang (✉) · F. Liu · C. Shen · F. Kong
Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei 230022, China
e-mail: zhangao@mail.ustc.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_71

the 1980s to detect bearing flaws. It uses wayside rail-mounted wayside monitoring microphones to collect the acoustic signal as the train passes by the detector [2–4]. However, there are some key techniques that need to be developed, one of which is the acoustic signals separation. For the acoustic signal acquired by a microphone is composed of a number of train bearing signals and noises, single signal of failure train bearing should be extracted to diagnose the fault type precisely in wayside fault diagnosis of train bearings.

There hasn't been any published study addressing on the acoustic signal separation in wayside fault diagnosis of train bearings. Blind Source Separation [5], a traditional signal separation method in fault diagnosis of rotating machinery, makes no use to the wayside train bearings acoustic signal separation, caused by the Doppler Effect and the seriously underdetermined problem. Signal separation is also a process of signal reconstruction for each acoustic source. However, the classical method, ridge extracting and reconstruction [6, 7], is also disabled, as the ridges extracted don't contain any physical information that we can distinguish which source they belong to. Meanwhile, ridges corresponding to different sources may mix together affecting by the Doppler Effect. Hence, a new method based on Dopplerlet transform [8], time-frequency domain filtering and inverse generalized S-transform [9] is proposed here to separate acoustic signals, and applied in the wayside fault diagnosis of train bearings.

2 Theory of the Acoustic Signal Separation Method

2.1 Doppler Effect

An assumption was made that an acoustic source moves with a constant speed along a straight railway. The scheme of the modeled measuring situation is presented in Fig. 1.

According to the acoustic theory of Morse [10], the sound intensity acquired by the microphone could be achieved, under the assumption that the acoustic source of the train bearing with subsonic velocity was a monopole point source and the medium was the ideal fluid. Moreover, the portion of high order with little effect

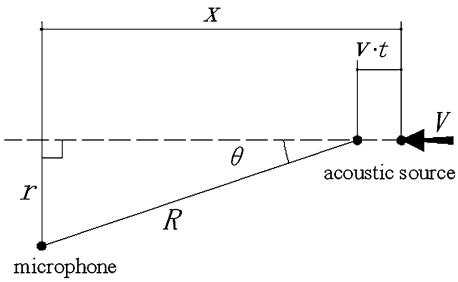


Fig. 1 Schematic diagram of the acoustic source with subsonic velocity

was ignored and the acoustic source is given as harmonic with the intensity of $q = q_0 \sin(\omega_0 t)$.

$$P = \frac{q_0 \omega_0 \cos(\omega_0(t - (R/c)))}{4\pi R(1 - M\cos\theta)^2} = \frac{q_0 \omega_0}{4\pi R(1 - M\cos\theta)^2} \times \cos(\omega_0(t - (R/c))) \quad (1)$$

where P represents the sound pressure acquired by the microphone, R represents the distance between the acoustic source and the microphone, θ denotes the angle between the connection from source to microphone and the direction of movement, c denote the sound velocity, v represents the moving speed of the source, and $M = v/c$ is the mach number of the source.

The amplitude of the signal acquired depends on the left side of the multiplication sign while the phase relies on the right side, with which the variation of the frequency could be attained.

$$f = f_0 \frac{M(x - vt) + \sqrt{(x - vt)^2 + (1 - M^2)r^2}}{(1 - M^2) \cdot \sqrt{(x - vt)^2 + (1 - M^2)r^2}} \quad (2)$$

where x represents the horizontal distance between the start point of the source and the microphone, r denotes the vertical distance between the motion trail and the microphone, $f_0 = \omega_0/2\pi$ is the stationary (base) frequency of the source, and f represents the instantaneous frequency received by the microphone.

2.2 Matching-Pursuits-Based Dopplerlet Transform (MPDT) for Signal Decomposition

The dilated and translated windowed Doppler functions (the Dopplerlets) is proposed by Zou and Li [8] to characterize both the linear and nonlinear time-frequency structures of a signal. However, the original Dopplerlet atom presented by Zou ignored the amplitude modulation and the time delay that sound travels from the source to the observer. So an improved Dopplerlet transform is proposed here to estimate the actual motion parameters.

2.2.1 Improved Gaussian Dopplerlet Atom

According to the theory acoustic theory of Doppler Effect in the above section, the improved Doppler signal (plural) is as follows

$$d'_{x,f_0,r,v,c}(t) = P \cdot \exp\{j2\pi f_0 \cdot (t - R(t)/c)\} \tag{3}$$

where $j = \sqrt{-1}$, and the variables are the same as those proposed in Sect. 2.1. Using this signal to modulate a normalized Gaussian window function, then dilating the modulated signal by a factor $\delta \in R^+$ yields

$$d_{x,f_0,\delta,r,v,c}(t) = (\pi\delta^2)^{-1/4} \exp\left\{-\frac{1}{2}\left(\frac{t - x/v}{\delta}\right)^2\right\} \cdot P \cdot \exp\{j2\pi f_0(t - R(t)/c)\} \tag{4}$$

which, for convenience, is called "Improved Gaussian Dopplerlet Atom" in accordance with the common convention in nomenclature in the signal-processing community.

2.2.2 Matching-Pursuits-Based Dopplerlet Transform

Once the Dopplerlets are used as atoms, the (complex) Dopplerlet transform of any square-integrable signal $s(t) \in L^2(R)$ may be readily defined as

$$\begin{aligned} DT_s(x, f_0, \delta, r, v, c) &= \langle s(t), d_{x,f_0,\delta,r,v,c}(t) \rangle \\ &= (\pi\delta^2)^{-1/4} \int_{-\infty}^{\infty} s(\tau) \exp^*\left\{-\frac{1}{2}\left(\frac{t - x/v}{\delta}\right)^2\right\} \cdot P \cdot \exp\{j2\pi f_0(\tau - R(\tau)/c)\} d\tau \end{aligned} \tag{5}$$

where $\langle , \rangle$ denotes the Dirac inner product and the superscript "*" denotes the complex conjugate operation.

Let $\gamma = (x, f_0, \delta, r, v, c)$ denote the parameter set, then the (complex) Dopplerlet transform can be written concisely in the form

$$DT_s(\gamma) = \langle s(t), d_\gamma(t) \rangle = \int_{-\infty}^{\infty} s(\tau) d^*_\gamma(\tau) d\tau \tag{6}$$

A matching pursuit algorithm [11, 12] is the one that adaptively decomposes any signal under analysis into a linear combination of a set of atoms that are selected from a large redundant dictionary of atoms in accordance with the criterion of maximum projection energy. These atoms are chosen in order to best match the signal's local structures.

Let H denote a Hilbert space, $D = \{d_\gamma\}_{\gamma \in \Gamma}$ be a dictionary of vectors in H with $\|d_\gamma\| = 1$ (where the index γ stands for the parameter set and Γ for the parameter space). Let $s(t) \in L^2(R)$; then, signal $s(t)$ can be decomposed as Eq. (7) with matching-pursuits-based Dopplerlet transform.

$$\begin{aligned} s &= R_s^{(0)} \\ &= R_s^{(1)} + \left\langle R_s^{(0)}, d_{\gamma 0} \right\rangle d_{\gamma 0} \\ &\vdots \\ &= R_s^{(k)} + \sum_{i=0}^{k-1} \left\langle R_s^{(i)}, d_{\gamma i} \right\rangle d_{\gamma i} \\ &\vdots \\ &= \sum_{i=0}^{+\infty} \left\langle R_s^{(i)}, d_{\gamma i} \right\rangle d_{\gamma i} = \sum_{i=0}^{+\infty} DT_s(\gamma_i) \cdot d_{\gamma i} \end{aligned} \tag{7}$$

where $R_s^{(i)}$ is its corresponding residual signal of each decomposition, and $\lim_{k \to +\infty} \left\| R_s^{(k+1)} \right\|^2 = 0$.

2.3 Generalized S-Transform (GST) and Inverse Generalized S-Transform (IGST)

The S-transform [13] is a time-frequency spectral localization method, similar to the short-time Fourier transform (STFT), but with a Gaussian window whose width scales inversely, and whose height scales linearly, with the frequency. The expression of the S-transform given by Stockwell is

$$S(\tau, f) = \int_{-\infty}^{\infty} h(t) \{ \frac{|f|}{\sqrt{2\pi}} \times \exp\left[\frac{-f^2(\tau - t)^2}{2} \right] \exp(-2\pi i f t) \} dt \tag{8}$$

where, S denotes the S-transform of h, which is a continuous function of time t; frequency is denoted by f; and the quantity τ is a parameter which controls the position of the Gaussian window on the t-axis. The scaling property of the Gaussian window is reminiscent of the scaling property of continuous wavelets [14] because one wavelength of the Fourier frequency is always equal to one standard deviation of the window.

The generalized S-transform (GST) [9] is obtained from the original S-transform, Eq. (8), by replacing the Gaussian window with a generalized window, denoted ω:

$$S(\tau, f, p) = \int_{-\infty}^{\infty} h(t) \omega(\tau - t, f, p) \exp(-2\pi i f t) dt \tag{9}$$

In Eq. (11), a set of parameters that govern the shape of ω are collectively denoted p. In practice, ω is replaced with a specific window, and p is replaced with an explicit parameter list enclosed in braces in subsequent usage. These parameters, along with S and ω, are assigned an identifying suffix. As an example, the Gaussian window, as modified by Mansinha, is denoted ω_{GS} and has the functional form

$$\omega_{GS}(\tau - t, f, \{\gamma_{GS}\}) = \frac{|f|}{\sqrt{2\pi}\gamma_{GS}} \exp\left[\frac{-f^2(\tau - t)^2}{2\gamma_{GS}^2}\right] \tag{10}$$

Here, γ_{GS} is the only element of p.

Hence, the inverse transformation of generalized S-transform (IGST) can be expressed as:

$$h(t) = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} S(\tau, f, p) d\tau \right\} \exp(2\pi i f t) df \tag{11}$$

2.4 Time-Frequency Filtering and Signal Reconstruction

According to the location of Dopplerlet atoms belong to its corresponding source in the time- frequency domain of GST acquired by MPDT, a time-frequency filter is designed to extract the component of signal that we interested in from the original signal. The time-frequency filter $H_\zeta(\tau, f)$ can be defined as

$$H_\zeta(\tau, f) = \begin{cases} 1 & (\tau, f) \in \zeta \\ 0 & (\tau, f) \notin \zeta \end{cases} \tag{12}$$

where, ζ is the location of interested signal component in time-frequency domain.

Therefore the component of signal that we interested in can be deduced through Eq. (13) as

$$h_\zeta(t) = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} S(\tau, f, p) H_\zeta(\tau, f) d\tau \right\} \exp(2\pi j f t) df \tag{13}$$

3 Procedure of Acoustic Signal Separation and Fault Diagnosis

Acoustic signals separation is a process of signal reconstruction for each acoustic source. The key techniques have been represented in the section mentioned above, while the processing steps are as follows.

(1) Search the parameters space to find the k primary functions-Dopplerlet atoms which match the original signal $s(t)$ best via MPDT, meanwhile, the parameter

sets $\{x, f_0, \delta, r, v, c\}$ compose those best primary functions could also be acquired.

(2) Judging from the parameter x, these Dopplerlet atoms are divided into different groups corresponding to diverse acoustic sources.
(3) Design a time-frequency filter through Eq. (12) to extract the data of Dopplerlet atoms corresponding to an acoustic source we interested in and its neighborhood in time-frequency domain.
(4) According to Eq. (13), reconstruct the signal of the acoustic source corresponding to a train bearing with the IGST.
(5) Resample the reconstructed signal via the method proposed by Dybala [15], and we will get the corrected signal $y(t)$ without Doppler distortion.
(6) Analysis the envelope spectrum of corrected signal $y(t)$ to diagnosis the fault type of the original signal $s(t)$.

4 Applied in the Wayside Acoustic Signal

To demonstrate the effectiveness of the proposed acoustic signals separation method for the wayside acoustic signals, an experiment was implemented with three speakers fixed on a vehicle which is moving at a constant speed of 110 km/h. The parameters of these loudspeakers are shown in Table 1. The instrumentations used in the experiment are as below: Microphone, type 4944-A, B&K; Data acquisition card, type PXI-4472, NI; PXI chassis, type PXI-1033, NI.

The multiple acoustic sources signal acquired by the Microphone is given in Fig. 2a which shows that the three types of signals are mixed together. According to the GST of the multiple acoustic sources signal shown in Fig. 2b, the phenomenon of Doppler distortion is obvious. A large number of Dopplerlet atoms are obtained via MPDT. Judging from the parameter x, the main Dopplerlet atoms of S1 are distinguished. Hence, a time-frequency filter is designed to extract the data of Dopplerlet atoms corresponding to S1 and its neighborhood in time-frequency domain. The GST result of the multiple acoustic sources signal after filtering is shown in Fig. 3a, which represents the main components of S1. Then, the recon-

Table 1 The parameters of loudspeakers

Acoustic source	x(m)	r(m)	Sound type
Speaker 1 (S1)	5	2	Outer race fault
Speaker 2 (S2)	6.5	2	Double-frequency interference signal (1400HZ, 2200HZ)
Speaker 3 (S3)	8	2	Inner race fault

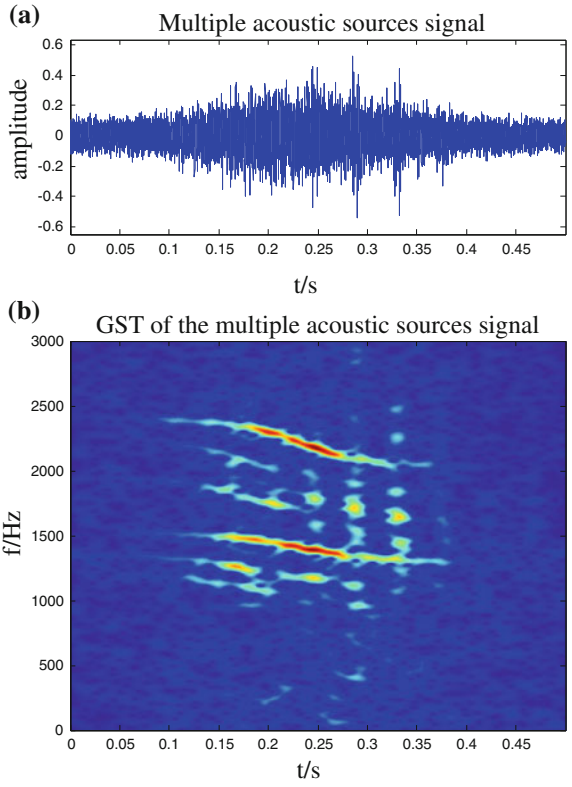


Fig. 2 Multiple acoustic sources signal in time domain and time-frequency domain

structed signal of S1 obtained via IGST is given in Fig. 3b. After re-sampling, the Doppler distortion effect of the reconstructed signal is eliminated, which could be seen clearly in Fig. 3c. Finally, the envelop spectrum shown in Fig. 3d is carried out, and the result f_o = 138.1 Hz is very approximate to 138.4 Hz, its theoretical value of outer race fault, which demonstrates that the reconstructed signal is indeed the outer race fault signal separated from the multiple acoustic sources signal.

Similarly, the process of separating S3 from the multiple acoustic sources signal is shown in Fig. 4a, b, c. In Fig. 4d, the inner race fault frequency f_i = 188.7 Hz, which is modulated by the rotational frequency f_r = 22.89 Hz, is also very approximate to its theoretical value 188.9 Hz. The result also demonstrates that the reconstructed signal is indeed the inner race fault signal separated from the multiple acoustic sources signal. Therefore, the method proposed in this chapter is effective and feasible.

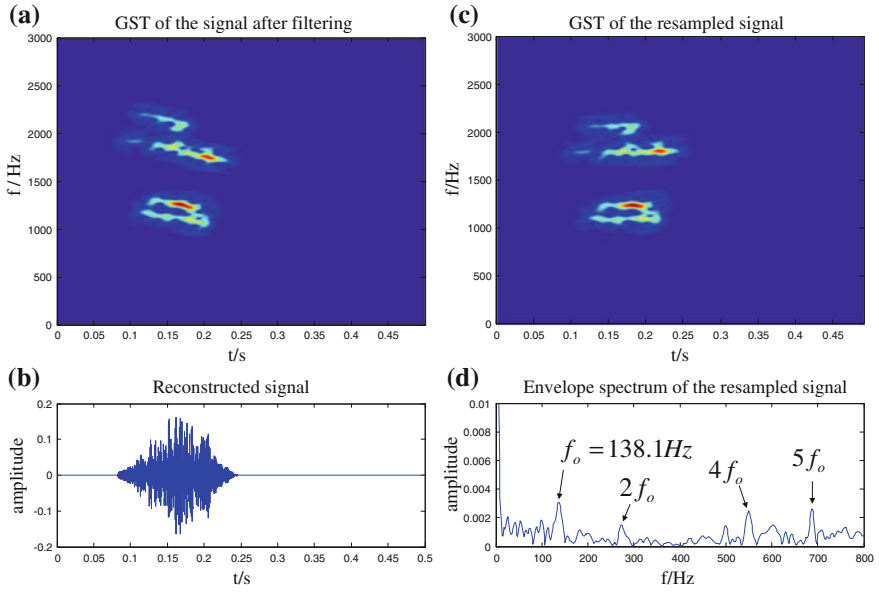


Fig. 3 The results of outer race fault signal reconstruction steps

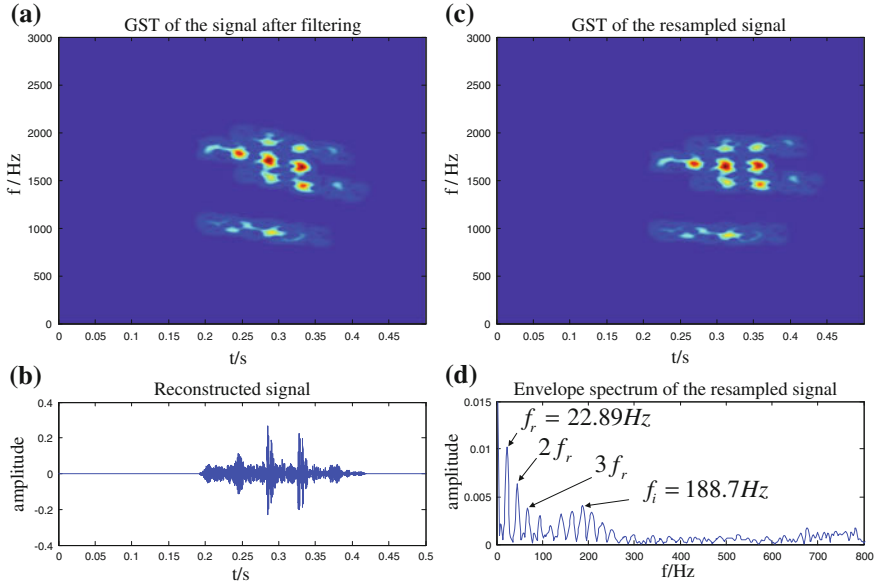


Fig. 4 The results of inner race fault signal reconstruction steps

5 Conclusion

Up to now, there hasn't been any published study addressing on the acoustic signal separation in wayside fault diagnosis of train bearings. In this chapter, a new method based on the Dopplerlet transform, time-frequency filtering and inverse generalized S-transform is proposed to separate different fault types of train bearing signals from the acoustic signal. An experiment with practical acoustic signals of train bearings with a defect on the outer race and the inner race is carried out in the end of this chapter, and the results verified the effectiveness of this method. Hence, this work could provide a technical reference for the wayside fault diagnosis of train bearings.

Acknowledgements This work is supported by the National Natural Science Foundation of China under Grant 51075379 and 51005221.

References

1. Choe HC, Wan Y, Chan AK (1997) Neural pattern identification of railroad wheel-bearing faults from audible acoustic signals: comparison of FFT, CWT and DWT features. SPIE Proc Wavelet Appl 3087:480–496
2. Irani FD (2002) Development and Deployment of Advanced Wayside Condition Monitoring Systems. Foreign Rolling Stock 39(2):39–45
3. Barke D, Chiu WK (2005) Structural health monitoring in the railway industry: a review. Struct Health Monit 4:81–93
4. Cline JE, Bilodeau JR (1998) Acoustic wayside identification of freight car roller bearing detects. In: Proceedings of the 1998 ASME/IEEE joint railroad conference, pp 79–83
5. Jutten C, Herault J (1991) Blind separation of sources, an adaptive algorithm based on neuromimetic architecture. Signal Process 24(1):1–10
6. Carmona R, Hwang WL, Torresani B (1999) Multiridge detection and time-frequency reconstruction. IEEE Trans Signal Process 47:480–492
7. Carmona R, Hwang WL, Torresani B (1997) Characterization of signals by the ridges of their wavelet transforms. IEEE Trans Signal Process 45:2586–2590
8. Zou H, Chen Y, Zhu J, Dai Q, Wu G, Li Y (2004) Steady-motion-based Dopplerlet transform: application to the estimation of range and speed of a moving sound source. IEEE J Oceanic Eng 29(3):887–905
9. Pinnegar CR, Mansinha L (2003) The S-transform with windows of arbitrary and varying window. Geophysics 68:381–385
10. Morse PM, Ingard KU (1986) Theoretical acoustics (Section 2), (Yang, X., et al, translate) Science Press, Beijing, pp 822–850
11. Mallat S, Zhang Z (1993) Matching pursuits with time-frequency dictio-naries. IEEE Trans Signal Processing 41:3397–3415
12. Papandreou-Suppappola A, Suppappola SB Adaptive time-frequency representations for multiple structures. In: Proceedings of the 10th IEEE workshop statistical signal and array processing, Pocono Manor, PA, Aug 2000, pp 579–583
13. Stockwell RG, Mansinha L, Lowe RP (1996) Localization of the complex spectrum: the S transform. IEEE Trans Signal Process 44:998–1001
14. Mallat S (1998) A wavelet tour of signal processing, Academic Press
15. Dybala J, Radkowski S (2012) Reduction of Doppler effect for the needs of wayside condition monitoring system of railway vehicles. Mech Syst Signal Process, in press

Fault Detection and Diagnostics Using Data Mining

Sun Chung and Dukki Chung

Abstract The purpose of data mining is to find new knowledge from databases in which complexity or the amount of data has so far been prohibitively large for human observation alone. Self-Organizing Map (SOM) is a special type of Artificial Neural Networks (ANNs) used in clustering, visualization and abstraction. In modern process automation systems, it is possible to collect and store huge amounts of measurement data. In this paper, SOM is used successfully to discover the base models from the automation system. Strategies based on data mining techniques are further developed for efficient fault detection and diagnostics. A semi-supervised anomaly detection technique is used with classification rules based on standardized data and domain experts' analysis to construct the condition monitoring system.

Keywords Data mining Self-Organizing map · Fault detection · Equipment diagnostics · Decision tree

1 Introduction

In modern industry, equipment maintenance is an important factor to ensure constant production. Manufacturing enterprises are facing increased maintenance, repair, and operation (MRO) costs at their plants. They normally operate an Asset Management system for the management of MRO. In order to reduce the MRO costs, process and equipment performance need to be analysed in further detail. Establishing efficient systems for fault diagnosis and condition monitoring of machines is an important part of maintenance policies. In this paper data mining

S. Chung (✉)
Cleveland State University, Cleveland, USA
e-mail: sschung@cis.csuohio.edu

D. Chung
Rockwell Automation, Milwaukee, USA
e-mail: dchung@ra.rockwell.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_72

techniques are used to provide efficient asset management decision support information derived from operations and maintenance data sources.

The purpose of data mining is to find new knowledge from databases in which complexity or the amount of data has so far been prohibitively large for human observation alone, and for which little prior understanding exists. Artificial Neural Networks (ANNs) offer tremendous opportunities for performing data mining activities, in particular problems pertaining to data classification and clustering. ANNs learn how to solve problems from data as opposed to solving problems based on problem specification. The Self-Organizing Map (SOM) is a special type of ANN used in clustering, visualization and abstraction. In modern process automation systems, it is possible to collect and store huge amounts of measurement data. In this paper, SOM is demonstrated successfully as a useful data mining tool for the discovery of models from large data sets produced by actual manufacturing lines in various plants.

The base model discovered by SOM is further enhanced by using a data mining technique based on a decision tree. This semi-supervised approach using domain experts' analysis in the second phase is used to construct the condition monitoring system.

2 Base Model Discovery Using Self-Organizing Map

Modern process automation systems daily produce huge amounts of data from various sensors and many devices attached to the control systems. For example, a particular plant line of interest produces about 1 million data points with many sensor measurements daily. It is therefore necessary to automate the examination of these data points to look for faults or failures. To filter out normal data points and ease the task of finding more important data, e.g., process anomalies, upsets or faults, a Self-Organizing Map (SOM) [1] is used.

SOM provides a way of representing multidimensional data in lower dimensional spaces—two dimensions in this implementation. The process of reducing the dimensionality of vectors is essentially a data compression technique known as vector quantization. SOM creates a network structure that preserves topological relationships between the training samples. A typical application of SOM is analysing complex vector data such as process states where data elements may be related to each other in a highly nonlinear fashion.

A SOM consists of nodes or neurons. Each node is associated with a weight vector of the same dimension as the input data vectors and a position in the map space. The usual arrangement of nodes is a regular spacing in a hexagonal or rectangular grid. The SOM describes a mapping from a higher dimensional input space to a lower dimensional map space. The procedure for placing a vector from input data space onto the map is to find the node with the closest weight vector to the vector taken from input data space and to assign the map coordinates of this node to the vector presented. A typical rectangular topology of SOM is presented in Fig. 1.

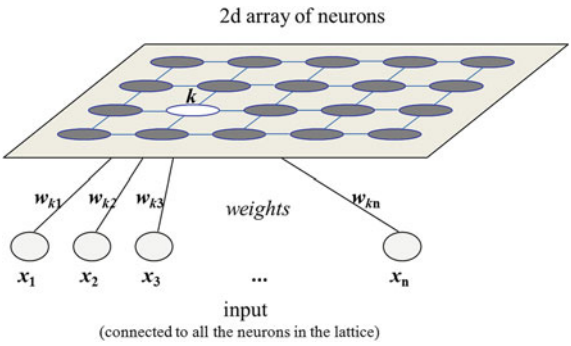


Fig. 1 SOM (*rectangular topology*)

Like most artificial neural network, SOM operates in two modes: learning and mapping. Learning builds the map using input samples. Mapping classifies a new input sample. Learning causes different parts of the network to respond similarly to certain input patterns. The training utilizes competitive learning. When a training example is fed to the network, its distance to all weight vectors is computed. The node with a weight vector most similar to the input is called the best matching unit. The weights of the best matching unit and nodes close to it in the SOM lattice are adjusted towards the input vector. This process is repeated for each input vector for a usually large number of cycles or epochs. The network ends up associating output nodes with groups or patterns in the input data set.

During mapping, there will be one single winning node: the node whose weight vector lies closest to the input vector. This can be simply determined by calculating the distance between the input vector and the weight vector.

One of the most important aspects of SOM is unsupervised learning. A SOM learns to classify the training data without any external supervision. Once the training samples are selected randomly from the collected data, the SOM learning process handles these data without any human intervention. After the learning process is finished, the input vectors will be mapped into specific regions of the constructed two dimensional map.

2.1 Training Algorithm

The network is created from a 2D lattice of nodes or neurons, each of which is fully connected to the input layer. Each node has a specific topological position (an x, y coordinate in the lattice) and contains a vector of weights of the same dimension as the input vectors.

Each weight vector w is updated by the following process

$$w_i(t+1) = w_{i(t)} + h_{c(x),i}(x(t) - w_i(t)) \tag{1}$$

where t is the sample index of the learning step, and where the learning step is performed recursively for each presentation of a sample of x. Index c is defined by the condition

$$\|x(t) - w_c(t)\| \leq \|x(t) - w_i(t)\| \quad \forall i \tag{2}$$

Here $h_{c(x),i}$ is called the neighbourhood function. It is a decreasing function of the distance between the i-th and c-th nodes on the lattice grid. As a typical example, the Gaussian neighbourhood function is defined by

$$h_{c(x),i} = \alpha(t) \exp\left(-\frac{\|r_i - r_c\|^2}{2\sigma^2(t)}\right) \tag{3}$$

where $0 < \alpha(t) < 1$ is the learning rate which decreases monotonically with the learning steps, $r_i \in \Re^2$ and $r_c \in \Re^2$ are the locations in the lattice, and $\sigma(t)$ corresponds to the width of the neighbourhood which is also decreasing monotonically with the learning step.

2.2 Initial SOM Mapping

SOM is a powerful algorithm for data visualization and dimension reduction, and can also be very effective for analysing sensor data from complex industrial processes. In this paper, SOM is used to build a base model of the automation process. The actual process of interest produces continuous sensor readings from 32 different sensors in every 0.1 s. A typical setup for the acquisition of sensor data from the automation process is shown in Fig. 2. Due to sheer volume of data, it is very difficult for a human to process and inspect all the sensor data at this rate.

To train SOM, 10,000 samples of sensor data are used. Each sample in the training set consists of floating point measurements from 32 different sensors. The training set consists of roughly 17 min of continuous sensor readings from the process.

A six by six rectangular lattice is used with a Gaussian neighbourhood function in the training phase. After training, SOM produced the map in Fig. 3. This map shows how many samples are mapped into each unit. Brighter shades mean more samples are mapped into that particular unit. By examining the mapping further, it is determined that only 28 samples out of 10,000 training samples are mapped into the two units in solid black. The remainder of the training samples are spread among the other units.

After further investigation, two of the sensors from those 28 samples in the training set exhibited unusual behaviours time to time, as shown in Fig. 4. These patterns suggest certain fault conditions in the equipment or the process.

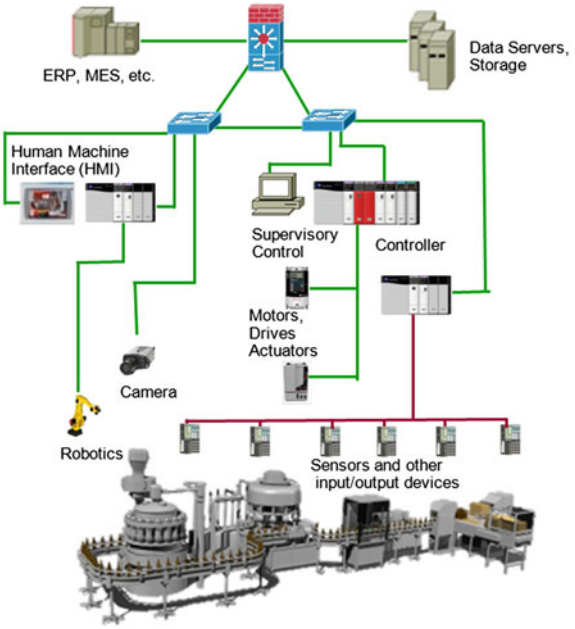


Fig. 2 Typical automation process line configuration for sensor data acquisition

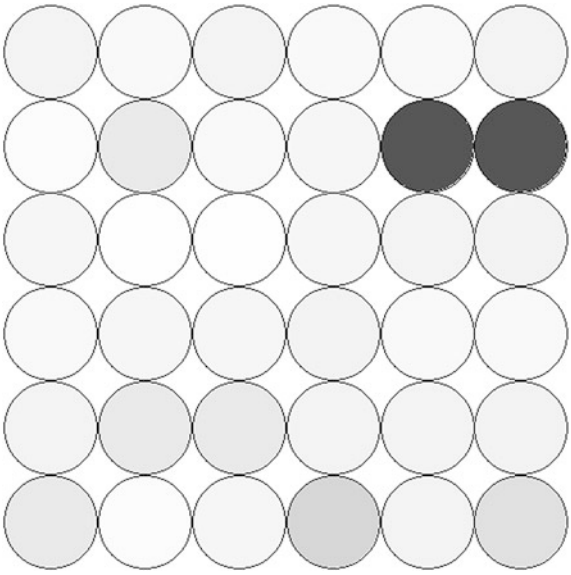


Fig. 3 SOM Mapping

Once the base model is constructed, it could be further enhanced with domain knowledge from human experts. SOM organizes the sample data so that the samples are usually surrounded by similar samples. However, similar samples are not

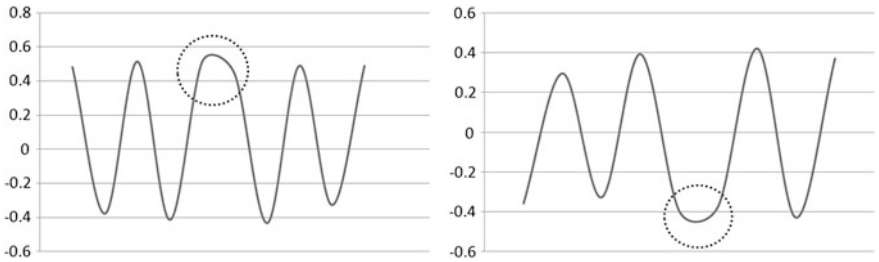


Fig. 4 Examples of unusual sensor patterns

always near each other. Sometimes clusters get split into smaller clusters. By incorporating domain knowledge from human experts, these shortcomings can be overcome. The approach used in this paper is further explained in Sect. 3.

3 Data Mining Based on Decision Tree Algorithm

Data mining techniques are feasible not only for database analysis but also for machine learning where high volumes of sensor data are prohibitively large for human observation alone. Fault diagnostics is based on pattern identification and classification. A data mining technology is introduced as a classification method for fault diagnostics of machinery. A method based on a decision tree algorithm and techniques to obtain a refined tree model that fits the machine fault data is developed in this section.

Among several types of classifiers, rule-based classifiers have a distinct advantage of being easily interpreted. This is a critical advantage especially in a data mining context where the overwhelmingly high volume of data often means that very little is known in advance about the underlying meaning of data and the mechanism which generates the data. Decision trees are the most popular form of rule-based classifiers and prediction tools since they have good performance and the rules generated from them are easily interpreted. The most widely used decision tree algorithms are C4.5 [2] and its most recent version C5.0 that came out in 2012 [3].

In this paper, SOM provides an initial method to evaluate and recognize the occurrence of machine faults. However, the knowledge is hidden in the network, so the rules cannot be easily extracted and interpreted. Initial classification performed by SOM is used by the decision trees algorithm as target classes. Decision trees can be more effectively applied to machine fault diagnosis since not only pattern classification, but also rule extraction and knowledge interpretation are required.

3.1 Decision Tree Approach

The Decision Tree algorithm is a classification and regression algorithm for use in predictive modeling of both discrete and continuous attributes. For discrete attributes, the algorithm makes predictions based on the relationships between input variables in a dataset. It uses the values, known as states, of those columns to predict the states of a target that is designated as predictable. Specifically, the algorithm identifies the input columns (variables) that are correlated with the predictable column (targets). The Decision Tree algorithm builds a data mining model by creating a series of splits in the tree. These splits are represented as nodes. The algorithm adds a node to the model every time that an input column is found to be significantly correlated with the predictable column. The way that the algorithm determines a split is different depending on whether it is predicting a continuous column or a discrete column.

The Tree Induction algorithm is based on a greedy strategy that splits the records of a node (an attribute) based on the attribute test that maximizes information gain on a target for prediction. For each distinct value (class) of an attribute, a frequency for each class in the dataset is counted to measure the information gain on a node using the count matrix to make decisions for the best node split. To determine the best split for each node, the information gain for each node split is measured by weighing each partition of the node with a scoring function. Entropy, Gini, and Bayesian Dirichlet Equivalence (BDE) are the most widely used scoring functions. C4.5 uses Entropy. The greedy approach prefers nodes with homogeneous class distribution and larger and purer partitions.

Entropy is represented as measure of uncertainty. Entropy at a given node t is defined as [4]:

$$Entropy(t) = -\sum_{\mathrm{j}} \mathrm{P(j|t)log2\ P(j|t)} \tag{4}$$

Note that $p(j|t)$ is the relative frequency of class j at node t. Maximum is log n_c when records are equally distributed among all classes implying least information. Minimum is 0.0 when all records belong to one class, implying most information gain. The decision for best node splitting is based on Information *GAIN* as follows [4]:

$$GAIN_{split} = Entropy(p) - \left(\sum_{i=1}^{k} \frac{n_i}{n} Entropy(i) \right) \tag{5}$$

where parent node p is split into k partitions; n_i is the number of records in partition i. GAIN measures reduction in entropy achieved because of the split. The algorithm chooses the split that achieves most reduction, thus maximizes *GAIN*. The algorithm tends to prefer splits that result in large numbers of partitions, each being small but pure. This disadvantage is adjusted by penalizing higher entropy partitioning, that is,

large numbers of small partitions. To overcome the disadvantage of entropy, the algorithm can choose to use a different weighing function. Bayesian Dirichlet Equivalence (BDE) is one of the more effective functions. Suppose that $\rho(\Theta_D|G)$ Dirichlet with equivalent sample size N' for some complete directed acyclic graph G in D, Then, for any Bayesian network B in D,

$$P(B,T) = P(B) \times \prod_{i=1}^{n} \prod_{j=1}^{q_i} \left(\frac{\Gamma(N'_{ij})}{\Gamma(N_{ij} + N'_{ij})} \right) \times \prod_{k=1}^{r_i} \left(\frac{\Gamma(N_{ijk} + N'_{ijk})}{\Gamma(N'_{ijk})} \right) \quad (6)$$

where $N'_{ijk} = N' \times P\left(X_i = x_{ik}, \prod X_i = w_{ij}|G\right)$. Equation (6) induces the likelihood-equivalence Bayesian Dirichlet (BDe) score [5].

A common problem in data mining models is that the model becomes too sensitive to small differences in the training data, in which case it said to be over-fitted or over-trained. An over-fitted model cannot be generalized to other data sets. To avoid over-fitting on any particular set of data, the Decision Tree algorithm uses techniques for controlling the growth of the tree. To stop the algorithm before it becomes a fully-grown tree, the following typical early stopping conditions for a node are applied for pre-pruning:

- Stop if all instances belong to the same class
- Stop if all the attribute values are the same

More restrictive conditions can be applied, if desired, as follows:

- Stop if number of instances is less than some user-specified threshold as a minimum support
- Stop if class distribution of instances are independent of the available features (e.g., using χ 2 test)
- Stop if expanding the current node does not improve information gain measures.

Some decision tree algorithms provide a choice of multiple scoring functions to find a better tree model for given data characteristics. This system uses a Microsoft Decision Trees algorithm that is enhanced from C4.5 and provides a choice of several scoring functions. There are various parameters which can be customized to get the best fitted tree model for a given data. COMPLEXITY_PENALTY controls the growth of the decision tree by penalizing too many tree splits, MINIMUM_SUPPORT is a threshold for the minimum number of leaf cases that is required to generate a split, and SCORE_METHOD chooses a different scoring function that weights node splits with a different model. These parameters can be used to control under-fitted and over-fitted trees. All the parameters and their meaning are listed in their technical references [6].

3.2 Initial Classification of a Base Model by SOM

Fault diagnostics is based on pattern identification and classification. However, it may be difficult to relate a fault to the data directly because the system is very complex and affected by numerous process parameters. One feasible method is to construct a feature-system state relationship using expert intelligence for reasoning and decision making to define normal states and various fault states. In this paper, 10,000 sensor data sets were collected with 32 different attributes (sensor values), and each continuous attribute was normalized and converted to 10 different states with a field expert's domain knowledge. In the first phase, SOM classifies the data into 36 different classes. Each class represents various machine states from normal states to fault states. These initial 36 classes are discovered by the SOM base model described in Sect. 2. By examining these 36 classes, the fault states are identified to be class 29 and class 30. There were total 28 cases out of 10,000 data sets that were categorized into class 29 and class 30. Our interest is to predict the fault classes—class 29 and class 30 out of 36 classes for the target. Of the 10,000 data sets, 8,000 were used for training and 2,000 were used as a test to predict the fault classes.

The greedy approach of decision tree algorithm with entropy as a scoring method tends to generate an over-fitting tree with pure but small partitions. To adjust the tree over-fitting problem, the Complexity Penalty parameter is used when measuring the purity of each split from a node. Microsoft Decision Trees has four scoring methods including the BDE function. Switching to the BDE scoring method adjusted the over-fitting tree problem in the initial tree with entropy as shown in Fig. 5. Figure 6 shows a tree generated with the BDE function which has fewer splits. This tree also has one of the most significant rules to identify (class 29) in the top level as shown in Fig. 7. Another way to avoid the problem is to use a scoring method other than entropy. Switching to the BDE scoring method adjusted the over-fitting tree problem. In addition, in order to prevent trees from over-fitting the data, all decision trees use some form of post pruning. Traditional pruning algorithms are based on error estimations—a node is pruned if the predicted error rate is decreased. However, this pruning technique does not always perform well, especially on imbalanced data sets, which is the case here. Chawla has shown that pruning in C4.5 can have a detrimental effect on learning from imbalanced data [7].

3.3 Optimization of Decision Tree Model on Imbalanced Data

The success of both decision trees and associate classifiers depends on the assumption that there is an equal amount of information for each class contained in the training data. In binary classification problems, if there is a similar number of instances for both positive and negative classes, C4.5 generally performs well. On the other hand, if the training data set tends to have an imbalanced class

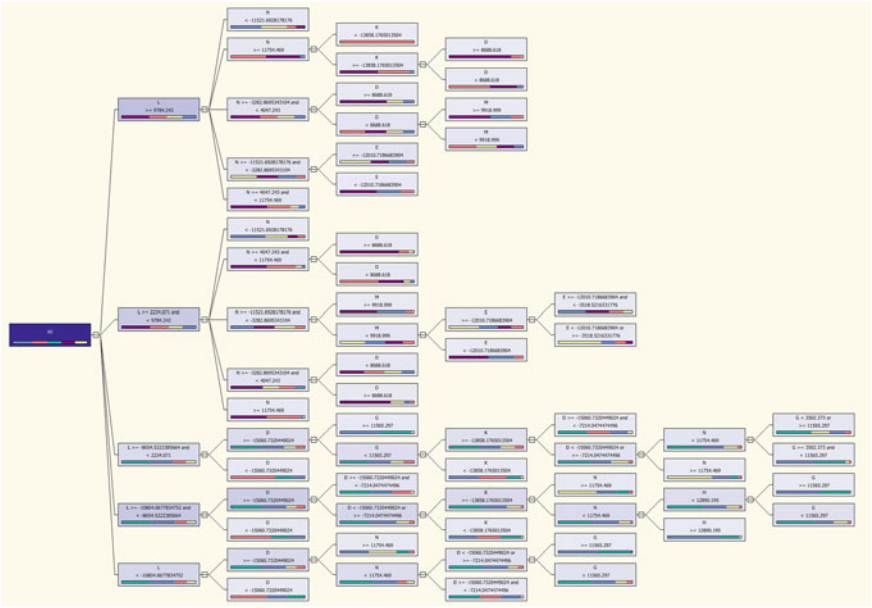


Fig. 5 Initial over-fitting decision tree using entropy

distribution, C4.5 will have a bias towards the majority class [8]. As it happens in our data, where the fault classes are a small fraction of the total (28/10,000), a prediction is typically related to the minority classes—the classes that are usually of greater interest.

The performance of Decision Tree Classifiers depends on the quality of the rules it discovers during the training process. In an imbalanced setting, Confidence is biased towards the majority class. Support and Confidence suggests that selecting the highest confidence rules means choosing the most frequent class among all the instances that contains that antecedent. However, for imbalanced data sets, since the size of the positive class is always much smaller than the negative class. Because of its low confidence, a "good" rule may be ranked behind other rules which have a higher confidence since they predict the majority class. For an imbalanced data set, high confidence rules do not necessarily imply high significance in imbalanced data, and some significant rules may not yield high confidence. This is a fatal problem with using a decision tree algorithm to classify and predict fault diagnosis, since it is often the minority class which is of more interest.

One way of solving the imbalanced class problem is to modify the class distributions in the training data by over-sampling the minority class or under-sampling the majority class. For instance, [9] uses over-sampling, by creating synthetic samples, to increase the number of minority class instances. Further variations on SMOTE [10] have integrated boosting with sampling strategies to

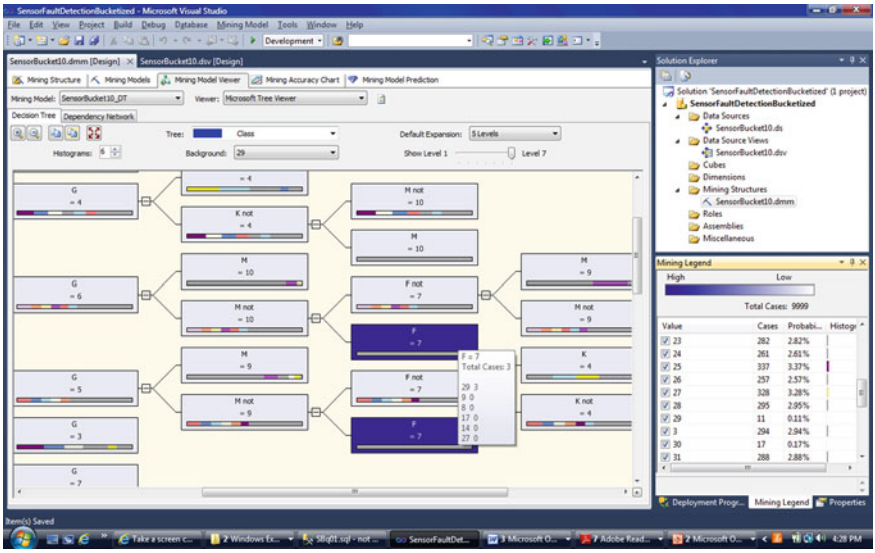


Fig. 6 Decision tree using BDE showing a rule for fault class 29

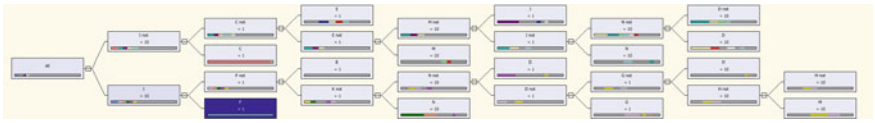


Fig. 7 Decision tree using BDE showing less splits

better model the minority class, by focusing on difficult samples that belong to both minority and majority classes. Data sampling is not the only way to deal with class imbalance problems. Cieslak and Chawla [11] uses the Hellinger Distance (HDDT) based on likelihood difference as the decision tree splitting criterion that was shown to be insensitive towards class distribution skewness. In [8], a new measure, Class Confidence Proportion (CCP), was proposed to classify imbalanced data sets.

To resolve the problem caused by imbalanced data, SMOTE [9] was adopted to generate over-sampling to increase the number of minority class instances by creating synthetic samples, in this case, for classes 29 and 30.

Applying a decision tree algorithm to this resampled data generated a best tree model that has very useful rules for two fault classes. Figure 8 shows the decision trees for the fault states 29 and 30. A much simpler and concise tree is generated, and the two most significant rules to detect classes 29 and 30 were shown on the top level of the trees as indicated by the dark nodes. The test results show excellent performance to predict these two fault classes.

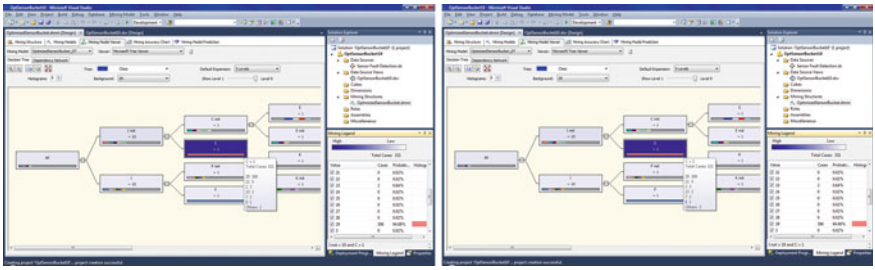


Fig. 8 Decision trees using SMOTE for fault class 29 and 30

4 Experiment Result

The same 10,000 sensor data sets previously mentioned in the paper from the automation process line were used for the experiment. The Microsoft Decision Trees algorithm was used with the training set to classify 36 states of the process line and to identify 20 explicit rules for machine status. The fault diagnostics system using the proposed methodology successfully diagnosed the real fault conditions. A node is pruned if the predicted error rate is decreased. Pessimistic error is used to validate the data set to estimate generalization error.

7398 data sets were generated after applying the SMOTE [9] sampling method to the original 10,000 data sets. Among 7398 data sets, 341 data sets and 357 data sets fall into class 29 and class 30 respectively. 30 % of 7,398 data sets were used for test. The misclassification rate for the training was overall 0.14 %. As shown in Fig. 3-2, the experimental results show that the system identifies 99.8 % of class 30 and 97.4 % of class 29 for the training set. On a separate test set, it achieves an overall accuracy of 100 % to predict the fault classes in the test data set: all the cases of the fault states are predicted correctly. Figure 9 shows the rules generated by the decision tree and the lift chart with the lift from 50 % by Random Guess Model to 100 % by the Optimized Decision Tree Model using the SMOTE method for target prediction.

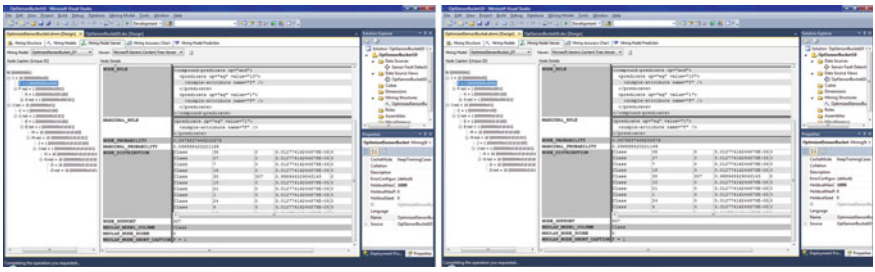


Fig. 9 Rules for fault class 30 and lift chart on decision Tree

The results indicate that data mining techniques can be effectively applied to diagnose an automation process line by useful rules to interpret the data generated by a decision tree algorithm. In addition, a data warehouse is built with the regularly generated sensor database and the results of the decision tree for further fault analysis. A time dimension and a location dimension are created in CUBE so that OLAP queries can be written to track down faults with the date and location of each occurrence. This work is in progress; the detail procedures and results may be published later.

5 Conclusion

The paper shows an application of a data mining technique to classify sensor data from an automation process line for fault diagnostics. SOM is used to discover a base model of the underlying process. A decision tree algorithm is used to refine the base model for classification and prediction. The paper presents problems and a methodology to generate a better fitted tree model for a given data and useful rules that are robust against class imbalance in the data set. The test results indicate that the proposed approach can be effectively applied to machine fault diagnostics.

References

1. Kohonen T (1995) Self-organizing maps. Series in information sciences, vol 30, 2dn edn. Springer, Heidelberg (1997)
2. Quinlan, JR (1993) C4.5: Programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco
3. Fakhr M, Elsayad AM (2012) Steel plates faults diagnosis with data mining models. J Comput Sci 8(4):506–514
4. Pang-Ning T, Michael S, Vipin K (2005) Introduction to data mining, p 769, 1st edn. May, 2005. ISBN-10: 0321321367 • ISBN-13: 9780321321367©2006. Addison-Wesley, Cloth. Published 05/02/2005
5. Alexandra MC (2009) Scoring functions for learning Bayesian networks. INESC-ID Tec. Rep. 54/2009 Apr 2009 IST, TULisbon/INESC-ID
6. Microsoft Decision Trees Algorithm technical references at http://msdn.microsoft.com/en-us/library/cc645868.aspx
7. Chawla NV (2003) C4. 5 and imbalanced data sets: investigating the effect of sampling method, probabilistic estimate, and decision tree structure
8. Liu W, Chawla S, Cieslaky Nitesh DA, Chawlay V (2010) A robust decision tree algorithm for imbalanced data sets. In: the Proceedings of SIAM International Conference on Data Mining
9. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16(1):321–357
10. Chawla NV, Lazarevic A, Hall LO, Bowyer KW (2003) SMOTEBoost: improving prediction of the minority class in boosting. Lecture notes in computer science, pp 107–119
11. Cieslak DA, Chawla NV (2008) Learning decision trees for unbalanced data. In: Proceedings of the 2008 European Conference on Machine Learning and Knowledge Discovery in Databases-Part I, pp 241–256. Springer, Heidelberg

Fault Detection of Planetary Gearboxes Based on an Adaptive Ensemble Empirical Mode Decomposition

Yaguo Lei, Naipeng Li and Jing Lin

Abstract Planetary gearboxes are widely used in modern industry because of their advantages of large transmission ratio, strong load-bearing capacity, etc. Planetary gearboxes differ from fixed-axis gearboxes and exhibit unique behaviors, which increase the difficulty of fault detection. The vibration based signal processing technique is one of the principal tools for detecting gearbox faults. Empirical mode decomposition (EMD), as a time-frequency analysis technique, has been used to process nonlinear and non-stationary problems. But it has the shortcoming of mode mixing in decomposing signals. To overcome this shortcoming, ensemble empirical mode decomposition (EEMD) was proposed accordingly. EEMD can reduce the mode mixing to some extent. The performance of EEMD, however, depends on the parameters adopted in the EEMD algorithm. In current studies on EEMD, the parameters were generally selected artificially and subjectively. To solve the problem, a new adaptive ensemble empirical mode decomposition method is proposed in this chapter. In the method, the sifting number is adaptively selected and the amplitude of the added noise changes with the signal frequency during the decomposition process. Both simulations and a case of fault detection of a planetary gear demonstrate that the proposed method obtains the improved results compared with the original EEMD.

Keywords Planetary gearboxes · Adaptive ensemble empirical mode decomposition · Fault detection

Y. Lei (✉) · N. Li · J. Lin
State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China
e-mail: yaguolei@mail.xjtu.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_73

1 Introduction

Planetary gearboxes are widely used in modern industry due to their advantages of large transmission ratio, strong load-bearing capacity, etc. Planetary gearboxes significantly differ from fixed-axis gearboxes and exhibit unique behaviours [1]. For example, multiple planet gears meshing simultaneously with the sun gear and the ring gear, and a large number of synchronous components (gears or bearings) in close proximity will excite similar vibrations in planetary gearboxes. These vibrations with different meshing phases couple with each other; as a result, the vibrations caused by faults could be buried by other normal vibrations [2]. In addition, there are multiple and time-varying vibration transmission paths from gear meshing points to transducers, which are typically fixed on the housing of planetary gearboxes. These transmission paths may attenuate the vibration signal of faulty components through dissipation and interference effects [3]. In addition, torques or loads applied to the gearboxes may also add to the effects of nonlinear transmission paths. All these effects would weaken the fault characteristics hidden in complicated vibration signals and increase the difficulty of fault diagnosis.

The vibration based signal processing technique is one of the principal tools for diagnosing gearbox faults [4]. It is possible to extract fault characteristics from vibration signals by performing the signal processing techniques on the signals. Empirical mode decomposition (EMD), as a time-frequency analysis technique, has been developed to process nonlinear and non-stationary problems. It is based on the local characteristic time scales of a signal and could decompose the complicated signal into a set of complete and almost orthogonal components named intrinsic mode function (IMF) [5, 6]. The IMFs represent the natural oscillatory mode embedded in the signal and work as the basis functions, which are determined by the signal itself, rather than pre-determined kernels. Thus, it is a self-adaptive signal processing method that can be applied to nonlinear and non-stationary process perfectly. However, one of the major drawbacks of EMD is the mode mixing problem, which is defined as either a single IMF consisting of components of widely disparate scales, or a component of a similar scale residing in different IMFs.

To overcome the problem of mode mixing in EMD, ensemble empirical mode decomposition (EEMD), an improved method of EMD, is presented [7]. EEMD is a noise-assisted data analysis method and by adding finite white noise to the investigated signal, the EEMD method is supposed to eliminate the mode mixing problem. The performance of EEMD, however, depends on the parameters adopted in the EEMD algorithm, such as the sifting number, the amplitude of the added noise, etc. In most of the current studies on EEMD, such parameters were set as the same values in disparate scales of the signal during the decomposition process. However, according to our study, different frequency components have different sensitivity to the amplitude of the noise. In other words, if the amplitude of the added noise is too small (large), the low (high) frequency components may be decomposed well, but the high (low) frequency components will have the mode

mixing problem. As a result, the problem of mode mixing is not solved well and the performance of EEMD needs to be improved further.

In this chapter, a new adaptive ensemble empirical mode decomposition method is proposed, in which the sifting number is adaptively selected and the amplitude of the added noise changes with the signal frequency during the decomposition process. The remainder of this chapter is organized as follows. Section 2 briefly introduces the method of EEMD. Section 3 is dedicated to a description of the proposed adaptive EEMD method. Section 4 gives a simulation example to illustrate the method. Section 5 shows a planetary gearbox test rig, on which some experiments were conducted and vibration data was acquired. The experimental data is utilized to demonstrate the performance of the proposed method. Both the simulation and the experimental results show that the adaptive EEMD obtains the improved results compared with the original EEMD. Section 6 draws concluding remarks.

2 Ensemble Empirical Mode Decomposition

EEMD was developed to solve the problem of mode mixing of EMD. It is a noise-assisted data analysis method, which defines the true IMF components as the mean of an ensemble of trials. Each trial consists of the decomposition results of the signal plus a white noise of finite amplitude [7, 8].

The principle of the EEMD algorithm is as follows: the added white noise would populate the whole time-frequency space uniformly with the constituting components of different scales. When a signal is added to this uniformly distributed white noise background, the components in different scales of the signal are automatically projected onto proper scales of reference established by the white noise in the background. Because each of the noise-added decompositions consists of the signal and the added white noise, each individual trial may certainly produce a very noisy result. But the noise in each trial is different in separate trials. Thus it can be decreased or even completely cancelled out in the ensemble mean of enough trails. The ensemble mean is treated as the true answer because finally, the only persistent part is the signal as more and more trials are added in the ensemble.

Based on the principle and observations above, the EEMD algorithm can be given as follows [9].

1. Initialize the number of ensemble M, the amplitude of the added white noise, and $m = 1$.
2. Perform the mth trial on the signal added white noise.

   (a) Add a white noise series with the given amplitude to the investigated signal

$$x_m(t) = x(t) + n_m(t) \tag{1}$$

   where $n_m(t)$ indicates the mth added white noise series, and $x_m(t)$ represents the noise-added signal of the mth trial.

(b) Decompose the noise-added signal $x_m(t)$ into I IMFs $c_{i,m}(i = 1, 2, \ldots, I)$ using the EMD method, where $c_{i,m}$ denotes the ith IMF of the mth trial, and I is the number of IMFs.
(c) If $m < M$ then go to step (a) with $m = m + 1$. Repeat steps (a) and (b) again and again, but with different white noise series each time.

3. Calculate the ensemble mean c_i of the M trials for each IMF.

$$c_i = \frac{1}{M}\sum_{m=1}^{M} c_{i,m},\ i = 1, 2, \ldots, I,\ m = 1, 2, \ldots, M \tag{2}$$

4. Report the mean $c_i(i = 1, 2, \ldots, I)$ of each of the I IMFs as the final IMFs.

EEMD is an improved version of EMD and is supposed to eliminate the mode mixing problem. The improvement of EEMD, however, depends on the parameters adopted in the EEMD algorithm, such as the sifting number, the amplitude of the added noise, etc. If these parameters are changed, the decomposition result will change accordingly. In the process of EMD, high and low frequency components have different sensitivity to noise to be added in the investigated signal. However, the same noise amplitudes and sifting number are used to all components in the original EEMD method. Therefore, the problem of mode mixing is not solved well and the performance of EEMD needs to be improved further.

3 Adaptive Ensemble Empirical Mode Decomposition

In this section, an adaptive EEMD is proposed to improve the original EEMD in solving the problem of mode mixing. In this method, according to different sensitivity of high and low frequency components to noise, larger noise and more sifting number are adopted in extracting higher frequency IMFs, while smaller noise and less sifting number are used in extracting lower frequency IMFs. To satisfy the requirement of noise, different kinds of noise are tried. It is found that the noise having the amplitude changing as a sinusoidal relation with its frequency performs best. Thus, the noise whose amplitude changes as a sinusoidal relation with its frequency, instead of white noise, is constructed and utilized in the EEMD algorithm. The frequency spectrum of the constructed noise is shown in Fig. 1, in which f_s represents the sampling frequency and e denotes the amplitude at the highest frequency. The sifting number for each IMF is adaptively set following Eq. (4). Figure 2 gives the flow chart of the adaptive EEMD. It includes the following procedural steps.

1. Initialize the amplitude e of the highest frequency of the added noise, the number of ensemble M, generally $M = 100$ and $e = 0.2$. Let $m = 1$.

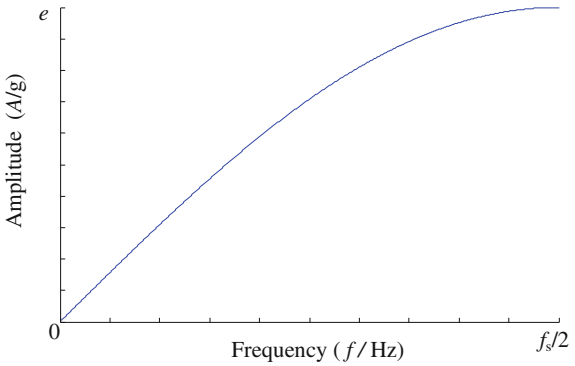


Fig. 1 The spectrum of the noise constructed

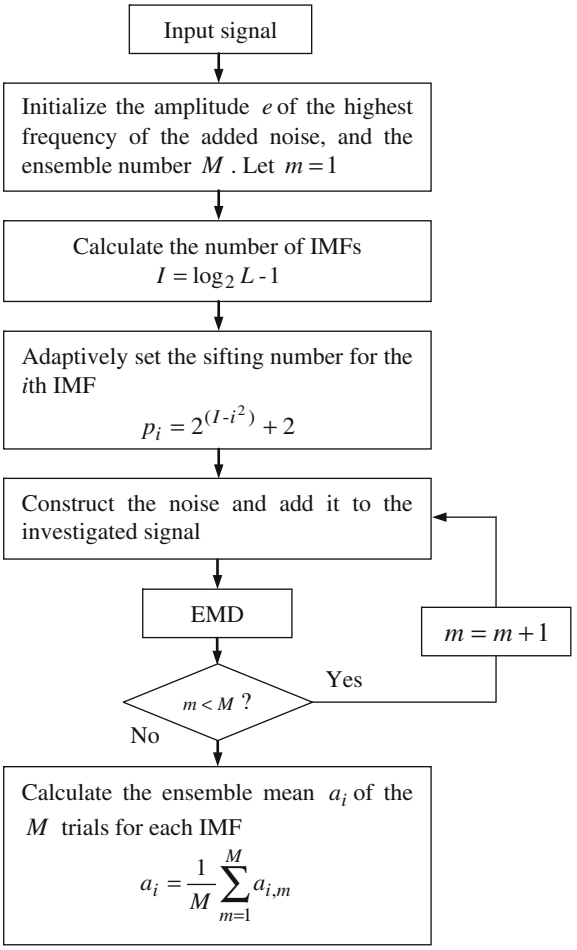


Fig. 2 Flow chart of the adaptive EEMD

2. Calculate the number of IMFs based on the signal length [7]

$$I = \log_2 L - 1 \tag{3}$$

where L is the signal length.

3. Adaptively set the sifting number p_i for the ith IMF according to the following equation.

$$p_i = 2^{(I-i^2)} + 2,\ i = 1, 2, \ldots, I \tag{4}$$

4. Construct the noise as shown in Fig. 1 and add it to the investigated signal.
5. Perform EMD on the signal added noise and obtain the mth decomposition result $a_{i,m}$.
6. If $m < M$ then go to step (4) with $m = m + 1$. Repeat steps (4) and (5).
7. Calculate the ensemble mean a_i of the M trials for each IMF and report the mean as the final IMFs.

$$a_i = \frac{1}{M}\sum_{m=1}^{M} a_{i,m},\ i = 1, 2, \ldots, I,\ m = 1, 2, \ldots, M \tag{5}$$

4 Simulation Experiment

In this section, a simulation signal is generated to illustrate the adaptive EEMD. Because modulation and impact are two typical fault events in mechanical fault diagnosis, the simulation signal includes modulation and impact components. The simulation signal also consists of two sinusoidal waves with a high and low frequencies respectively to represent certain rotating frequencies of machinery. Thus, there are altogether four components corresponding to different physical meaning in the simulation signal. The four components and the simulation signal combined by them are shown in Fig. 3a–e, respectively.

Applying the adaptive EEMD method to the decomposition of the simulation signal, the decomposed first four IMFs are plotted in Fig. 4. It can be seen from Fig. 4 that IMFs 1–4 respectively correspond to the impact component, the modulation component, the high-frequency sinusoidal wave and the low-frequency sinusoidal wave. Comparing the decomposed IMFs shown in Fig. 4 with the real components given in Fig. 3, it is found that the different components embedded in the signal can be extracted accurately using the adaptive EEMD.

For comparisons, the simulation signal is analyzed again using the EMD method and the original EEMD method. The decomposition results are displayed in Figs. 5 and 6, respectively. In Fig. 5, it is clear that the mode mixing is occurring in all of the first four IMFs decomposed by the EMD, and the EMD cannot decompose the four elements absolutely. The decomposed result in Fig. 6 is better than the above

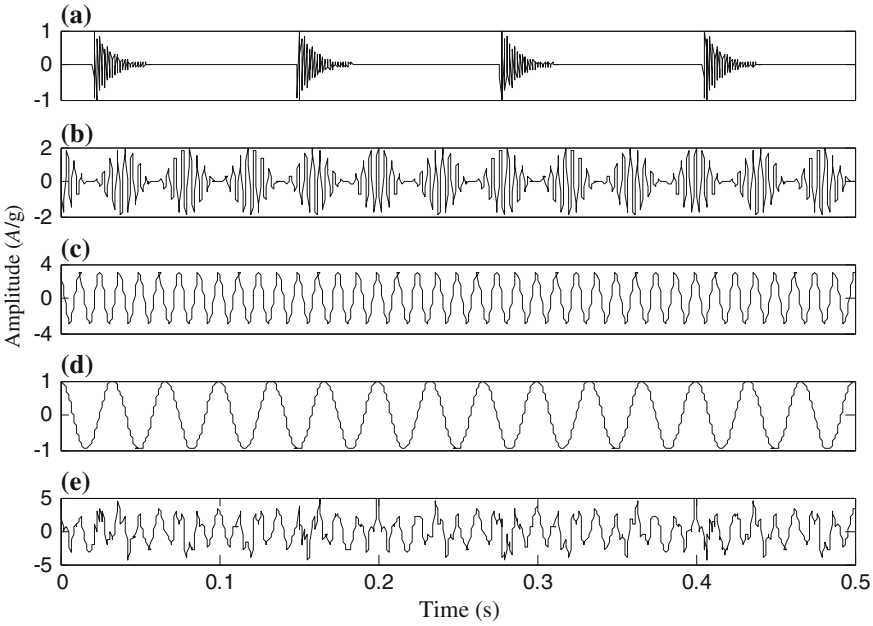


Fig. 3 **a**–**d** the four components, **e** the simulation signal

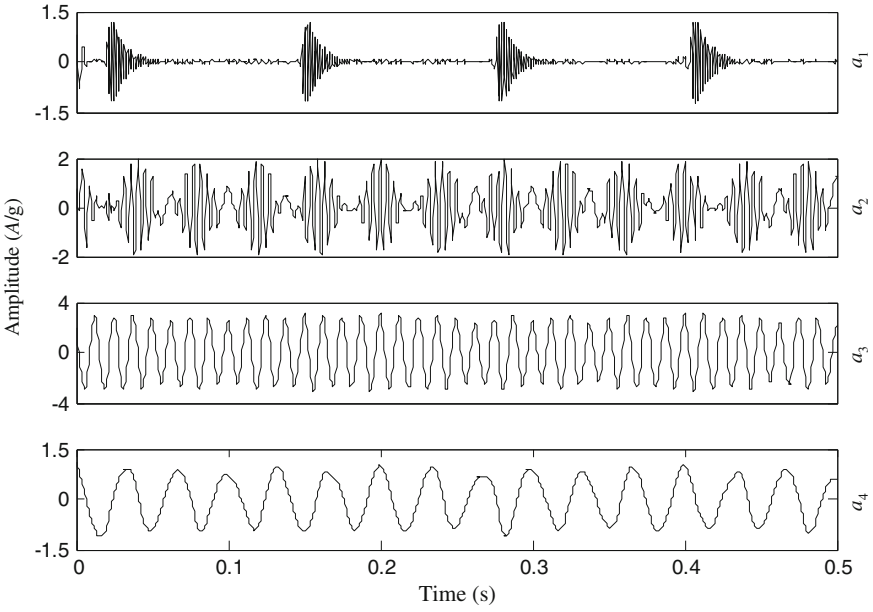


Fig. 4 The decomposed result of the simulation signal using the adaptive EEMD

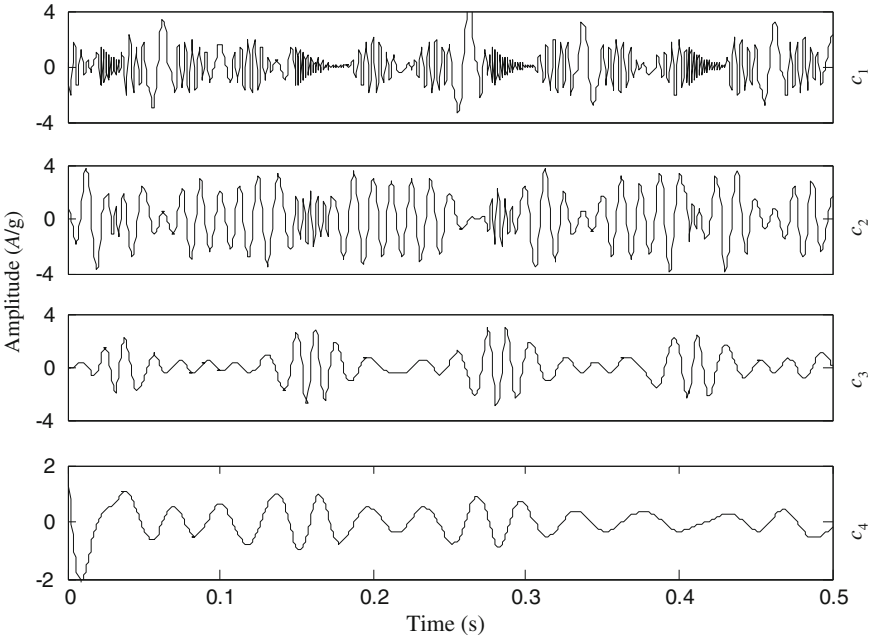


Fig. 5 The decomposed result of the simulation signal using the EMD

one, the high-frequency and the low-frequency sinusoidal waves are separately decomposed into the third and the forth IMFs. However, mode mixing still happens in the modulation and impact components which can be seen from the first IMF. What's more, the amplitude of the second IMF corresponding to the modulation component changes obviously. These results show that the original EEMD method can reduce the mode mixing to some extent, but can't decompose the two fault components into different IMFs exactly.

Based on the above simulation and comparisons, it is concluded that the adaptive EEMD solves the problem of mode mixing more effectively and produces more accurate IMFs than the original EEMD, by adding noise having the amplitude changing as a sinusoidal relation with its frequency into the signal, and adaptively changing the sifting number for different IMFs.

5 Fault Detection of Planetary Gearboxes

In order to demonstrate the effectiveness of the adaptive EEMD method in fault diagnosis of planetary gearboxes, a planetary gearbox test rig is used and experiments on it are conducted. The planetary gearbox test rig includes two gearboxes, a 3-hp motor for driving the gearboxes, and a magnetic brake for loading. The motor

rotating speed is controlled by a speed controller. The load is provided by the magnetic brake and can be adjusted by a brake controller. One gearbox in the test rig is a planetary one, the other is a fixed-axis one. The present study just concerns fault detection of the planetary gearbox, in which an inner sun gear is surrounded by multiple rotating planet gears, and a stationary outer ring gear [10]. A crack at the tooth root of one planetary gear is created in our experiments.

The motor speed is set about 20 Hz. The accelerometer is mounted on the planetary gearbox casing. An NI data acquisition system and a laptop with the data acquisition software are used to collect the vibration data for further processing. The sampling frequency is 5,120 Hz. Some parameters and the characteristic frequencies of the planetary gearbox are summarized in Table 1. From the table, it is observed that the rotating frequency of the planetary gear is 2.5 times as large as that of the carrier. Therefore, when the carrier rotates 2 cycles, the planetary gear meshes 5 periods with the ring gear, i.e. 200 teeth. This tooth number is twice as large as that of the ring gear. That is to say, the ring gear meshes 2 periods with the planetary gear. In other words, the planetary gear returns to the initial position whenever the carrier rotates 2 cycles. For the carrier to finish rotating 2 cycles, it takes 2/3.33 = 0.6 s.

The measured vibration signal from the test rig is illustrated in Fig. 7a. Figure 7b shows the frequency spectrum of the signal. It is observed that there are a series of impulses in the time-domain waveform. The period of the impulses is $t = 0.1$ s.

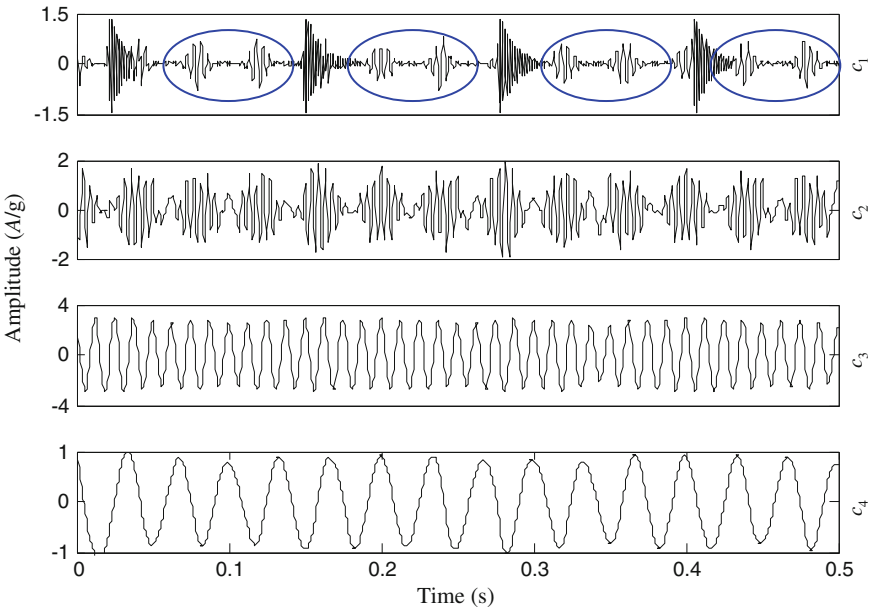


Fig. 6 The decomposed result of the simulation signal using the original EEMD

Table 1 Parameters of the planetary gearbox

The tooth number of sun gear	The tooth number of planetary gears	The tooth number of ring gear	The number of planetary gears	The rotating frequency of carrier/Hz	The rotating frequency of planetary gears/Hz	The mesh frequency/Hz
20	40	100	3	3.33	8.33	333.33

That means that the impulse frequency is 10 Hz. In the planetary gearbox of the test rig, there are three planetary gears. They pass the fixed transducers in turn, and therefore the pass frequency of the planetary gears equals 3 times as large as the rotating frequency of the carrier, i.e. 10 Hz. It is obvious that the impulses in the time-domain waveform are caused by the rotation of the carrier and belong to the vibration components of the normal gearbox. Besides these impulses, we do not find any fault characteristics. The reason is that the fault characteristics of the planetary gearbox are masked by the normal vibration components. We also investigate the frequency spectrum of the vibration signal in Fig. 6b. It is observed that there are rich sidebands around the mesh frequency and the interval of the sidebands is 3.33 Hz, equal to the rotating frequency of the carrier and not the fault characteristics. Thus, the fault characteristics of the planetary gear crack are not detected based on both the time-domain waveform and its frequency spectrum.

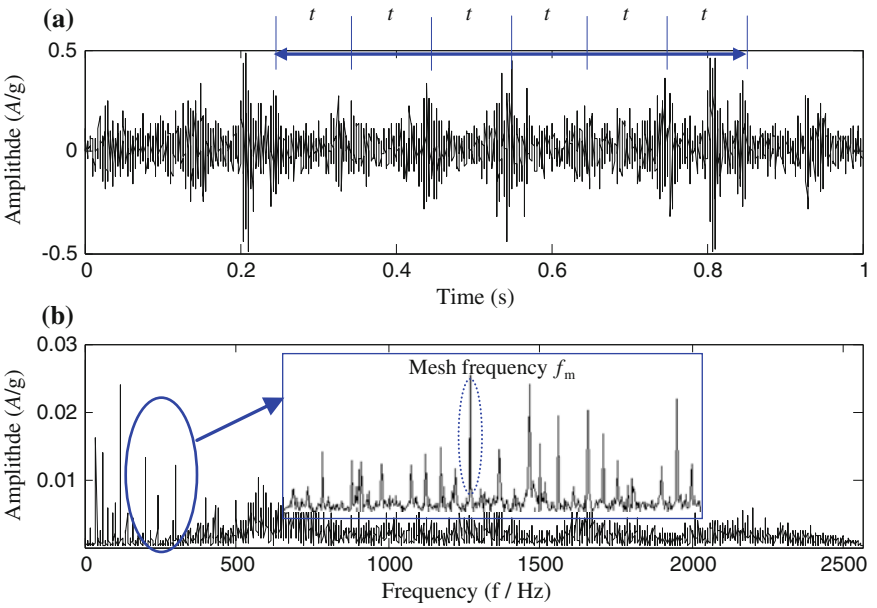


Fig. 7 The experimental vibration signal, **a** time-domain waveform, and **b** frequency spectrum

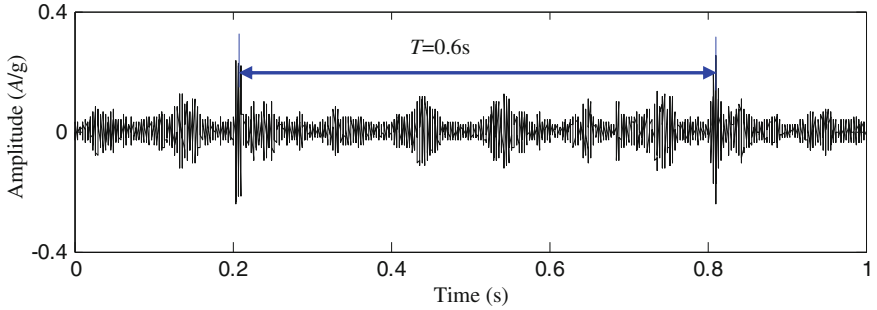


Fig. 8 The first IMF decomposed by the adaptive EEMD

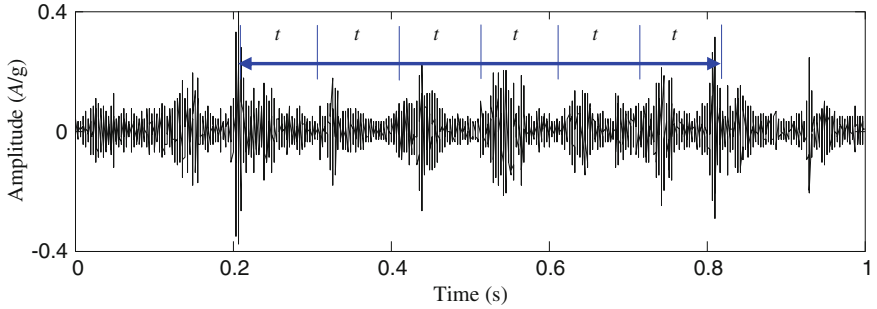


Fig. 9 The first IMF decomposed by the original EEMD

Then the proposed adaptive EEMD is used to process the above signal. The first IMF decomposed, given in Fig. 8, contains the richest information among all IMFs and therefore it is selected for further analysis. It is seen from the figure that there are impulses with the period $T = 0.6$s. Based on the above analysis, it is concluded that whenever the carrier rotates 2 cycles, the planetary gear returns to the initial position. Thus, the fault period of the damaged planetary gear is twice as large as the rotating period of the carrier, i.e. 0.6 s. So, the adaptive EEMD is able to extract the impulse component caused by the cracked planetary gear from the normal components effectively.

The original EEMD is also applied to analyse the same signal and the first IMF is displayed in Fig. 9. Although it is seen that there are periodic impulses in the waveform of the IMF, the impulse ($T = 0.6$s) caused by the cracked gear and those ($t = 0.1$s) caused by the rotation of the carrier are decomposed in the same IMF. That is to say, the mode mixing is occurring. Through the comparisons, it is drawn that the adaptive EEMD is more effective than the original EEMD in detecting faults of the planetary gearboxes.

6 Conclusion

In planetary gearbox transmissions, multiple meshing pairs of planet/sun and planet/ring produce similar vibrations and the measured vibration signals come from all of the interactions after they propagate through the complex transmission paths. Thus it is challenging to extract the fault characteristics of the planetary gearboxes. The vibration based signal processing technique is one of the useful tools for diagnosing gearbox faults. This chapter proposes an improved method named adaptive ensemble empirical mode decomposition (EEMD) for fault detection of planetary gearboxes. In the adaptive EEMD, the amplitude of the added noise changes with the signal frequency and the sifting number is adaptively selected during the decomposition process. Simulations are generated to compare the adaptive EEMD and the original EEMD. It is noticed that the former produces more accurate IMFs than the latter. Then, the method is applied to crack detection of a planetary gear in a test rig and it reveals clearer fault characteristics compared with the original EEMD.

Acknowledgements This research is supported by National Natural Science Foundation of China (51222503 and 51005172), New Century Excellent Talents in University (NCET-11-0421), Provincial Natural Science Foundation research project of Shaanxi (2013JQ7011), and Fundamental Research Funds for the Central Universities (2012jdgz01).

References

1. Lei YG, Han D, Lin J et al (2013) Planetary gearbox fault diagnosis using an adaptive stochastic resonance method. Mech Syst Signal Process 38:113–124
2. Blunt DM, Keller JA (2006) Detection of a fatigue crack in a UH-60A planet gear carrier using vibration analysis. Mech Syst Signal Process 20:2095–2111
3. Hines JA, Muench DS, Keller JA et al. (2005) Effects of time-synchronous averaging implementations on HUMS features for UH-60A planetary carrier cracking. American Helicopter Society 61st Annual Forum, Grapevine, TX, 1–3 June 2005
4. Lei YG, Lin J, He ZJ et al (2011) Application of an improved kurtogram method for fault diagnosis of rolling element bearings. Mech Syst Signal Process 25(5):1738–1749
5. Huang NE, Shen Z, Long SR et al (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non–stationary time series analysis. Proc Roy Soc London A 454:903–995
6. Lei YG, Lin J, He ZJ et al (2013) A review on empirical mode decomposition in fault diagnosis of rotating machinery. Mech Syst Signal Process 35(1–2):108–126
7. Wu ZH, Huang NE (2009) Ensemble empirical mode decomposition: a noise–assisted data analysis method. Adv Adapt Data Anal 1:1–41
8. Lei YG, He ZJ, Zi YY (2009) Application of the EEMD method to rotor fault diagnosis of rotating machinery. Mech Syst Signal Proc 23:1327–1338
9. Lei YG, Zuo MJ (2009) Fault diagnosis of rotating machinery using an improved HHT based on EEMD and sensitive IMFs. Meas Sci Technol 20:2280–2294
10. Lei YG, Lin J, He ZJ et al (2012) A method based on multi-sensor data fusion for fault detection of planetary gearboxes. Sensors 12:2005–2017

Building Diagnostic Techniques and Building Diagnosis: The Way Forward

A.K.H. Kwan and P.L. Ng

Abstract As buildings become old, their structural conditions deteriorate, causing concerns of irreparable damage and structural safety. To address these concerns of aged buildings, regular inspection and condition assessment for the purpose of building diagnosis are required. The inspection may consist of visual inspection, crack mapping, deflection measurement, settlement measurement, and observations of signs of water leakage and steel corrosion, whereas the condition assessment generally comprises of taking samples for materials testing, in situ measurement of temperature, moisture, half-cell electrical potential, vibration and delamination, and occasionally even continuous monitoring. However, in Hong Kong, not all of the test and measurement methods are accredited and often different laboratories/personnel follow different practices. Finally, building diagnosis has to be performed to make a judgment on the overall structural condition in terms of expected residual life and the repair needed. This requires good knowledge of structural engineering, materials and testing. Hence, building diagnosticians should be recognised as professionals of a special discipline, but this is not happening yet.

1 Introduction

There are lots of post-World War II buildings in Hong Kong that are already more than 50 or even 60 years old. Most of the public housing blocks more than 40 years old have been redeveloped and replaced by new ones, but many private buildings more than 50 years old are still around. Relatively, because of dispersed ownership and unwillingness of the owners to pay for maintenance, the conditions of private

A.K.H. Kwan
Department of Civil Engineering, The University of Hong Kong, Hong Kong, China
e-mail: khkwan@hku.hk

P.L. Ng (✉)
Department of Civil Engineering, The University of Hong Kong, Hong Kong, China
e-mail: irdngpl@gmail.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_74

buildings are usually worse than those of public housing blocks at the same age. The many dilapidated buildings in Hong Kong are giving people a shabby impression of the city, albeit modern.

Depending on the maintenance provided, the conditions of old buildings could vary widely and some old buildings are in very bad shape with obvious concrete cracking, concrete spalling, steel corrosion, water leakage and excessive deflection etc. In reality, the design life of common buildings is only 50 years and when buildings come close to the end of their respective expected life span, the buildings would start to have various health problems and in the worst case even terminal diseases. The only way to ensure good health of a building is to provide proper maintenance. Building maintenance, which could amount to more than HK $100,000 per repair per dwelling unit, is quite costly; nonetheless, this is imperative to the safety and serviceability of the building, and could reduce the rate of value depreciation of the premise.

Building maintenance is not just for the purpose of avoiding health deterioration of the buildings. In fact, it should also be viewed from public safety perspective. We could imagine the disastrous consequence resulted from a piece of concrete falling off from the wall of a multi-storey building onto a crowded street and hitting a vehicle or pedestrian in the street. Fortunately no one has been killed from incidents of such nature but this scenario is bound to happen sooner or later if we keep ignoring the maintenance of old buildings in Hong Kong.

Before we plan our maintenance and repair works, building inspection and condition assessment should be conducted for building diagnosis [10]. Building inspection is more on the overall and general conditions, as can be directly observed or measured. It may consist of visual inspection, crack mapping, deflection measurement, settlement measurement, and observations of signs of water leakage and steel corrosion. Condition assessment is more on detailed investigations and analysis. It generally comprises of taking samples for materials testing, in situ measurement of temperature, moisture, half-cell electrical potential, vibration and delamination, and occasionally continuous monitoring of movement and water leakage.

In theory, both building inspection and condition assessment should be entrusted to professionals with good knowledge and experience in materials and testing. However, in reality, this is not the case probably because there are insufficient professionals with adequate knowledge and experience. Moreover, the building diagnostic tests should all be accredited, but actually some of the building diagnostic tests are not yet accredited. Different personnel adopt different practices because there are no official guidelines established so far to regulate the performance of tests. Some equipment operators, technicians and report writers might not possess the expertise required. This situation has recently been steered to improve with the introduction of the Mandatory Building Inspection Scheme (MBIS) in 2012 [8]. The MBIS was launched following the enactment of amendments to the Building Ordinance and Building (Inspection and Repair) Regulation in 2011.

Under the MBIS, the building inspection shall be carried out by Registered Inspector (RI), who must be a learned person, as assessed and approved by the Buildings Department to perform building inspection works. In tandem with the MBIS, joint efforts need to be paid by the government, learned societies, and the practitioners.

At the outset, we need to bear in mind that building diagnostic testing and building diagnosis are not the same; the former is performing specific tests for obtaining data for interpretation while the latter is interpreting the data so obtained and making a judgment on the overall structural condition in terms of residual life and the repair needed. Building diagnosis, which requires structural safety appraisal, must be entrusted only to professionals with good structural sense.

2 Building Inspection

Visual inspection is the prime step of building inspection. Before the visual inspection, the building and structural plans, and the construction and maintenance records of the building should be obtained for preliminary study. During the visual inspection, particular attention shall be paid to additions and alterations (whether legal or illegal), the inspector should also identify the structural components and non-structural components, observe the presence of cracks, record any signs of water leakage and steel corrosion, tap at plasters, tiles and concrete surfaces to detect delamination, check the straightness of structural members to detect excessive deflection, and check the inclination of the building using a plumb line. All the observed defects should be marked on drawings for detailed desk top study together with the building and structural plans. At this stage, it may be necessary to liaise with the building owner for more information. In order to avoid missing out traits and information that are important to the ensuing investigations, the first visual inspection must be led by an experienced professional.

Following the first visual inspection and the desk top study, further field investigation is required including crack mapping, measurements of deflection, settlement and inclination, locating the possible sources of water leakage and a more thorough survey of the identified defects. Some non-destructive test methods may be used for a quick and preliminary appraisal. These include: covermeter survey of concrete cover to steel rebars, ultrasonic pulse velocity tests for detecting voids and defects in concrete, rebound hammer tests for rough estimation of concrete strength, impact echo test for detecting delamination, infrared thermography for remote detection of delamination and/or water leakage, and surface penetration radar for detecting internal cracks and defects etc. An account of non-destructive testing and evaluation of concrete structures was presented by Maierhofer et al. [14]. Brief description of these non-destructive test methods is given in the following.

2.1 Covermeter Survey

The working principle of covermeter is based on electromagnetism. Electric current in the coil winding of the probe generates a magnetic field which propagates through the concrete and interacts with any metal buried therein, such as reinforcing steel. The interaction causes a secondary magnetic field to propagate back to the probe where it is detected by another coil, or in some instruments by modifying the primary field. The signal received will increase with increasing rebar size and decrease with increasing rebar distance (concrete cover). By making certain assumptions about the rebar and specifically by assuming the presence of only one rebar within the primary magnetic field, the instrument can be calibrated to convert the intensity of signal to distance and hence the cover thickness. Reference is made to British Standard BS 1881 Part 204 [5] for the guidelines of conducting covermeter survey. However, if there is more than one rebar within the range of the primary field, the instrument will receive a greater signal and indicate a shallower cover than the true cover. Some manufacturers claim that the size of the reinforcing bar may be determined by the use of spacer blocks and associated in-built mathematical processing. Such methods work satisfactorily only where a single rebar is present within the range of the probe. Therefore, the accuracy of covermeter is mainly affected by grouped reinforcing bars of unknown bar sizes.

2.2 Ultrasonic Pulse Velocity Test

In the ultrasonic pulse velocity test, the time of travel of an ultrasonic pulse through the concrete structure is measured and the pulse velocity is determined by the relation: pulse velocity = distance/time. As void and defects in the concrete prevent direct passage of ultrasonic pulse owing to the existence of concrete-air interfaces, the ultrasonic test can reveal internal defects of concrete such as the presence of honeycombing at the interior. Besides, as there is positive relationship between wave velocity and elastic modulus, as well as between elastic modulus and strength, the ultrasonic velocity is able to reflect the concrete strength. Reference can be made to British Standard BS EN 12504 Part 4 [6] and American Standard ASTM C 597 [1] for the guidelines of conducting the ultrasonic pulse velocity test. The equipment and field work of the test are illustrated in Fig. 1.

2.3 Rebound Hammer Test

Rebound hammer test, or Schmidt hammer test, is a simple method to estimate the in situ concrete strength. Guidelines for conducting the rebound hammer test can be referred to British Standard BS EN 12504 Part 2 [7]. The hammer measures the

Fig. 1 Ultrasonic pulse velocity test

rebound of a spring loaded mass impacting against the surface of the concrete. The rebound hammer has an arbitrary scale ranging from 10 to 100. Empirical correlation was established between concrete strength and the rebound number. It should be noted that the surface for testing should be grinded flat and smooth. When conducting tests, the hammer should be held at right angles to the surface, because the rebound reading can be affected by the orientation of the hammer. When used on the underside of a suspended slab, gravity will increase the rebound distance of the mass (vice versa for a test conducted on the top surface of a floor slab). Each rebound hammer should be calibrated before use. The major drawback of rebound hammer test is the limited accuracy. Even for calibrated hammers, the error of test could be about 15 %; whereas for uncalibrated hammers, the accuracy is much worse and the error can reach 30 %.

2.4 Impact Echo Test

In the impact echo test, a short-duration mechanical impact, produced by tapping a small steel sphere against the concrete surface, generates low-frequency stress waves (up to about 80 kHz) that propagate through the structure and are reflected by flaws and/or external surfaces. Multiple reflections of these waves within the structure excite local modes of vibration, and the resulting surface displacements are recorded by a transducer located adjacent to the impact. The piezoelectric crystal in the transducer produces a voltage proportional to the displacement, and the resulting voltage-time signal (called a waveform) is digitized and transferred to the memory of a computer, where it is transformed mathematically into a spectrum of amplitude versus frequency. The dominant frequencies, which appear as peaks in the spectrum, are associated with multiple reflections of stress waves within the structure, or with flexural vibrations in thin or delaminated layers. The fundamental equation of impact-echo is: depth of flaw = wave speed/frequency/2. The frequency

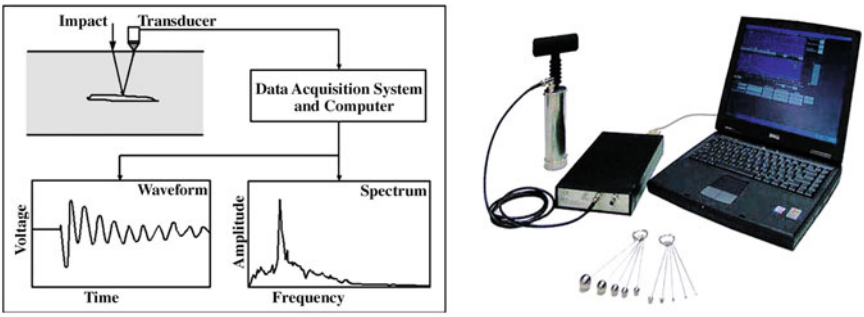


Fig. 2 Impact echo test

is obtained from the test as the dominant frequency of the signal, whereas the wave speed should be measured prior to the test. Guidelines of conducting the impact echo test can be referred to ASTM C 1383 [3]. Figure 2 depicts the schematic diagram of and equipment for impact echo test.

2.5 Infrared Thermography

The use of infrared thermography (or abbreviated as IRT) in structural damage assessment is one of the broad applications of thermal imaging. Thermographic camera detects radiation in the infrared range and produce images of the radiation. At the surface of concrete structure, regions with moisture trapping, water leakage, concrete spalling, debonding of tiles, etc. emit different amount of infrared radiation, and show up in the thermographic image by their different temperature transmittance. The procedures to conduct IRT can be referred to the specific test manual [11]. Examples of thermographic images are displayed in Fig. 3. The main features of IRT are: free of contact by remote sensing, full-field examination of large areas, poses no requirement of human accessibility, generation of real-time images for rapid detection, compatibility with digital post-processing, and ability of radiation to penetrate mist. On the other hand, the limitations of IRT include: (1) the test is qualitative rather than quantitative; (2) only the surface is measured but not the interior; (3) for delamination failure, the thickness of delamination cannot be assessed; (4) the surface temperature can be altered by human activities and climatic factors such as rain and wind; (5) the test is interfered by reflected solar radiation, external shadings, shadows cast by nearby structures, and radiation from surroundings; (6) thermal radiation is obstructed by the presence of objects between the camera and detected surface; (7) test results are affected by the thickness of rendering and services buried in the structure; (8) accuracy deteriorates with distance due to attenuation of thermal radiation; (9) viewing at large angle of elevation introduces distortion to the image; and (10) difficulty in the interpretation of test results arose from noise and variation in emissivity.

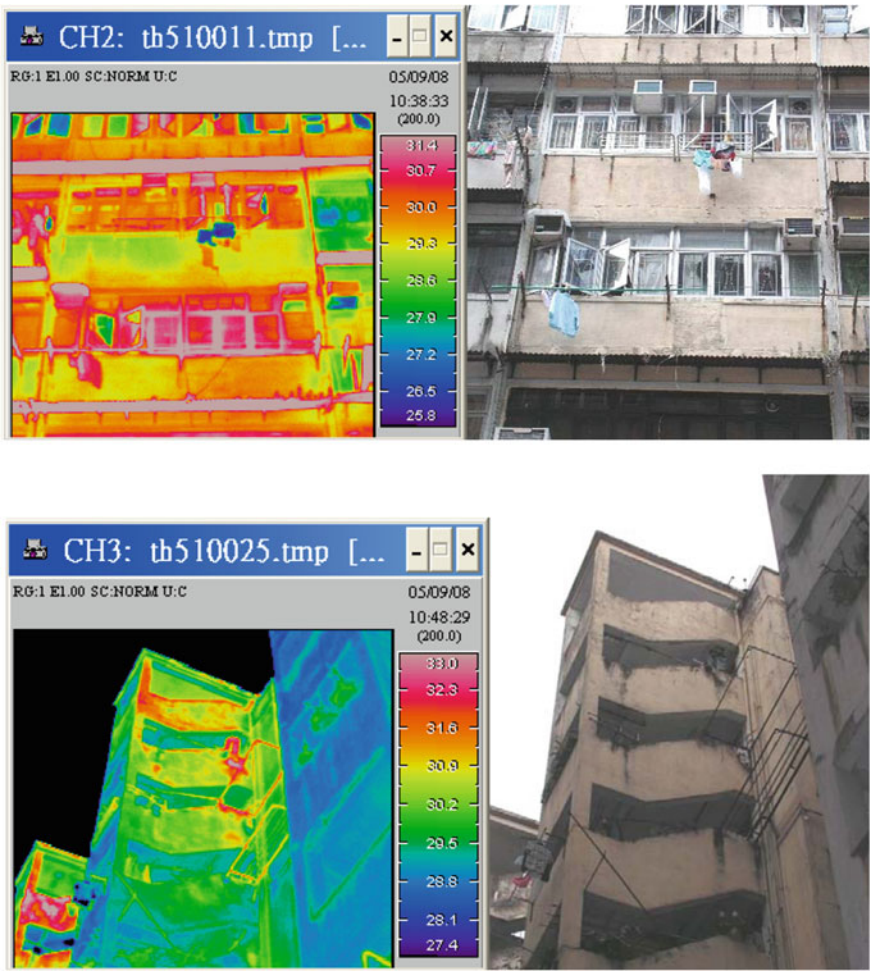


Fig. 3 Thermographic images from infrared thermography

2.6 Surface Penetration Radar

Surface penetration radar, or SPR, involves the propagation of pulses of electromagnetic waves in concrete structure, and these waves are reflected when they encounter a material that has substantially different electrical properties (or dielectric constants) from concrete. It allows determination of concrete cover, spatial distribution of steel reinforcement, location of cast-in objects, and detection of internal defects, where electrically contrasting layers exist (between concrete and steel, or between concrete and air) and partial reflection of incident energy occurs at the interface. The concrete cover or thickness can be determined via the propagation

velocity of the electromagnetic wave through the studied concrete, which is governed by the equation: propagation velocity through concrete = propagation velocity in free space (i.e. 3×10^8 m/s)/$\sqrt{}$dielectric constant. The test procedures using SPR are contained in the specific test manual [12]. Figure 4 illustrates the usage of SPR. SPR has the merits of high accuracy and reliability, and the results of SPR can be readily digitized and processed by visualization software to facilitate the tracking of defects. On the other hand, there are limitations of SPR, including limited depth of test concrete due to wave attenuation and dispersion, and inability to cope with relatively conductive test surfaces such as very wet or saturated concrete surface and concrete containing slag aggregate with high iron content.

The crack mapping and the non-destructive test results of structural components should be sent to concrete experts for detailed study. There are many possible causes of concrete cracking, such as plastic shrinkage, plastic settlement, early thermal movement, temperature variation, sulphate attack, acid attack, alkali-silica reaction, rebar corrosion, overloading, vibration and fire damage etc. and therefore rigorous analysis by a real concrete expert is needed, as misunderstandings in concrete cracking behaviour leading to false conclusions are commonplace in the industry. For example, some engineers simply attribute the water leakage through cracks in concrete to drying shrinkage of the concrete, without paying regard to the false logic in claiming the concrete to be drying while there is water leakage. Dependent on the situation, there may be a necessity to conduct some more tests during the condition assessment to find out the exact causes of the cracks, because the crack repair method is dependent on the causes of cracking.

The above-mentioned non-destructive tests should be carried out by an accredited laboratory (in Hong Kong, the accreditation protocol is the Hong Kong Laboratory Accreditation Scheme, or HOKLAS, operated by Hong Kong Accreditation Service, or HKAS), which has these tests accredited (note that HOKLAS does not just accredit a laboratory, it accredits also each individual test to be carried out by the laboratory). This is a point of importance as some professionals may not be fully aware of the operation mechanism of HKAS and HOKLAS, and they just accept test certificates issued by an HOKLAS accredited laboratory without verifying whether the laboratory has the specific tests accredited. Nevertheless, we do need to bear in

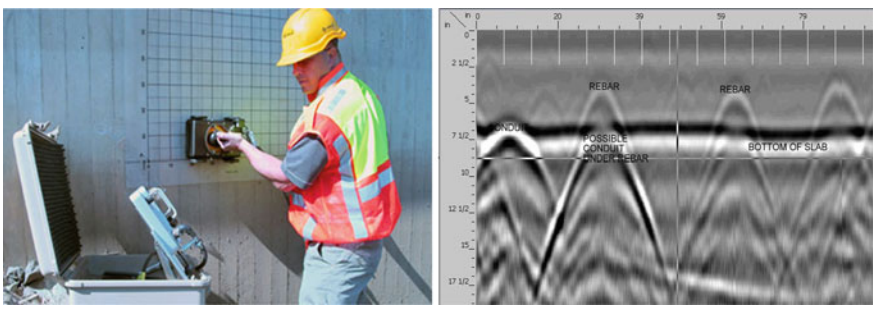


Fig. 4 Surface penetration radar survey

mind that some of the above-mentioned tests are quite new and have never been accredited at all. As the way forward, standardization and accreditation of each test are required.

3 Condition Assessment

Subsequent to the building inspection, a detailed plan for condition assessment should be worked out. At the minimum, core samples should be taken from the structural components (i.e. the walls, columns, beams and slabs) for concrete strength tests, dry powder samples should be taken for chloride content tests, and carbonation depth measurement should be carried out. Detailed description of the testing and assessment methodology can be found in Bungey et al. [9]. For an overall appraisal, the sampling locations should be representative of each environmental condition (internal and external, facing the sea and facing the hill, normally dry, normally wet and cyclically dry and wet etc.), each type of structural component (wall, column, beam and slab) and each grade of concrete.

Past experience revealed that the strength of concrete in some old buildings could be rather low and in extreme cases as low as only 5 MPa. When such low concrete strength is encountered, a full structural checking of the load carrying capacity of the building is required. Moreover, due probably to the use of sand containing salt as fine aggregate in the concrete and the use of seawater for flushing of toilets, the chloride content in the concrete could far exceed the permissible limits. For buildings more than 30 years old, there is also a high probability that the carbonation depth has reached beyond the embedded steel rebar surfaces. At locations with high chloride content or large carbonation depth, resistivity and half-cell electrical potential measurements should be carried out, as explained in the following.

3.1 Carbonation Test

The carbonation test is to determine the carbonation depth. A phenolphthalein solution is sprayed onto freshly exposed concrete surface (as phenolphthalein is insoluble in water, ethanol is employed as the solvent). The solution turns pink when pH > 8.6, and remains colourless when pH ≤ 8.6. The carbonation depth is measured as the average depth of the colourless region, in which the alkalinity had been neutralized by carbonation. Specification of the carbonation test was published by the Hong Kong Housing Authority [13]. It should be noted that as de-passivation of steel can take place at pH below 10.5, the carbonation test does not fully reflect the extent of possible steel corrosion. Figure 5 shows the colour change of phenolphthalein sprayed onto concrete.

Fig. 5 Carbonation test of concrete cores

3.2 Resistivity Measurement

As corrosion of steel is an electro-chemical process, the electrical resistance of the concrete will cast influence on the corrosion rate. The lower is the electrical resistance, the more readily the corrosion current flows through the concrete and the greater is the probability of corrosion. This property is utilized in the non-destructive testing of concrete structures by using a four-probe resistivity meter (Fig. 6). Among the four probes, the two outer probes pass a current, and the inner probes measure voltage difference. ASTM D 3633 has provided guidelines for resistivity measurement [4].

3.3 Half-Cell Electrical Potential Measurement

As mentioned in the above, the corrosion of steel in concrete is an electro-chemical process, similar to the reaction in a galvanic cell (i.e. a battery). The electro-chemical reaction produces an electric current, which is measurable as an electric

Fig. 6 Resistivity measurement

field on the concrete surface. This potential field is measured with an electrode known as half-cell, which is made up of a piece of metal in its own solution, e.g. copper (Cu) in copper(II) sulphate ($CuSO_4$) solution. By making measurements over the whole concrete surface, distinction between corroding and non-corroding locations can be identified; and by producing isopotential contour map of the surface, different zones of varying degrees of corrosion can be demarcated. Guidelines for conducting the half-cell electrical potential measurement can be referenced to ASTM C 876 [2]. Figure 7 depicts the schematic diagram and field work of half-cell electrical potential measurement. The drawbacks of using half-cell potentiometer are as follows. Firstly, it requires small open-up into the concrete member for the probe to be in contact with the embedded reinforcement. Secondly, surface preparation of concrete is required. Thirdly, the results are largely dependent on the effectiveness of electrical contact. Fourthly, the protective or decorative coatings applied to concrete may introduce errors to the results. Finally, the potentiometer makes no indication of the corrosion rate but only the probability that corrosion is underway.

Generally, where there are high chloride contents, large carbonation depths or signs of water leakage, resistivity measurement should be carried out (the resistivity reflects the moisture condition because wet concrete has low resistivity whereas dry concrete has high resistivity). High chloride content and/or large carbonation depth coupled with high water content (low resistivity) would lead to a high potential of steel corrosion. Half-cell electrical potential measurement may also help to identify potential areas of steel corrosion. At such areas, concrete repair to replace the concrete covers and restore passivity protection to the steel rebars, application of coating to stop further ingress of moisture, chloride and carbon dioxide, and injection of corrosion inhibitors may be advisable. In addition to the above, at locations where large temperature variations are expected such as on the roof or near the roof (related to thermal cracking), inside or outside cold storage (related to condensation of water droplets on concrete surfaces) and concrete surfaces subjected to strong sunshine (related to ageing of polymer and adhesives), measurement or even continuous monitoring of temperature using thermal couples should be carried out.

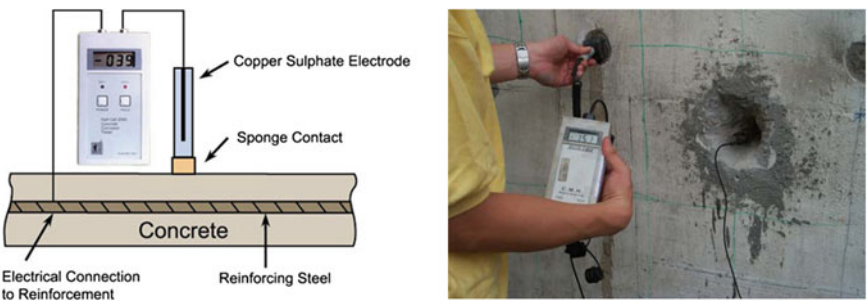


Fig. 7 Half-cell electrical potential measurement

Table 1 Test methods and reference standards for condition assessment

Test method	Reference standard
Chloride content	CS1: 2010 section 21
Core strength test	CS1: 2010 section 15
Surface penetration radar	HKCI: TM2
Half-cell potential	ASTM C 876
Impact echo	ASTM C 1383
Infrared thermography	HKCI: TM1
Alkali silica reactivity	ASTM C 856
Rapid chloride permeability	CS1: 2010 section 19
Resistivity	ASTM D 3633
Ultrasonic pulse velocity	BS EN 12504-4, ASTM C 597
Covermeter	BS 1881 Part 204
Phenolphthalein test (carbonation)	HKHA MTS specification
Rebound hammer	BS EN 12504-2
Tensile test of steel reinforcing bars	CS2: 2012
Tensile test of structural steelworks	BS 4360 or BS EN 10025

The various test methods and their reference standards are listed in Table 1. These test methods are based on the British Standards, the European Standards, the American Standards, the Hong Kong Construction Standards, and prevalent specifications and test method manuals in Hong Kong. Each test method has its own limitations and thus several tests may be required. As there is a lack of prevailing authoritative guidelines, the authors have come across some laboratories performing the above tests without following any recognised standards or seeking accreditation, and yet their reports were accepted. Moreover, the sampling rates and the acceptance criteria of some of these tests have remained a matter of engineering judgment, leading to widely different practices by different diagnosticians or laboratories. Apparently the development of universal guidelines to regulate the building diagnostic tests is needed.

4 Building Diagnosis

Building diagnosis is not the same as building diagnostic testing. Even with all the building diagnostic tests accredited and only well-trained or approved personnel allowed to perform the accredited tests, there are still problems of how the test results should be interpreted and how to make a judgment on the overall structural condition, the residual life or the probability of achieving the designed working life, and the repair needed. As an analogy, in medicine, diagnosis is the job of a medical doctor, whereas diagnostic testing is the job of a medical laboratory. In building inspection and maintenance, diagnosis is the job of professionals called building

diagnosticians, whereas diagnostic testing is the job of construction materials testing laboratories.

Building diagnosis is a specialty by itself and building diagnosticians should be recognised as professionals of a special discipline. This is not happening yet because many people just claim themselves to be building diagnosticians without making due regard to the high knowledge requirements of structural engineering, materials and testing. Building diagnosticians should be professionally qualified with good knowledge of structural engineering, materials and testing. Preferably, building diagnosticians should also be able to carry out forensic investigations on the probably causes of various defects in buildings, or alternatively, the building diagnostician could refer to specialists when necessary.

5 Conclusions

The Mandatory Building Inspection Scheme (MBIS) for buildings in Hong Kong has come into force, and in accordance with the scheme, the building inspection works are entrusted only to Registered Inspectors (RI), who must possess the necessary knowledge. Various test methods for building inspection and condition assessment have been presented in this chapter, including covermeter survey, ultrasonic pulse velocity test, rebound hammer test, impact echo test, infrared thermography, surface penetration radar, carbonation test, resistivity measurement, and half-cell electrical potential measurement. To ensure that the building diagnostic tests are properly and reliably conducted, the tests must be carried out by an accredited laboratory with the specific tests accredited. In this regard, the Hong Kong Accreditation Service (HKAS) plays the important role of master control and to set a reasonably high standard for accreditation. The Buildings Department also plays the important role to enforce the requirements of accreditation and to administer the MBIS. To enable the RIs to make good judgment as building diagnosticians, guidelines and training are needed.

References

1. ASTM International (2009) ASTM C 597–09: standard test method for pulse velocity through concrete. American Society for Testing and Materials, Pennsylvania
2. ASTM International (2009) ASTM C 876–09: standard test method for corrosion potentials of uncoated reinforcing steel in concrete. American Society for Testing and Materials, Pennsylvania
3. ASTM International (2010) ASTM C 1383–04(2010): standard test method for measuring the p-wave speed and the thickness of concrete plates using the impact-echo method. American Society for Testing and Materials, Pennsylvania
4. ASTM International (2012) ASTM D 3633–12: standard test method for electrical resistivity of membrane-pavement systems. American Society for Testing and Materials, Pennsylvania

5. British Standards Institution (1988) BS 1881: testing concrete: part 204: recommendations on the use of electromagnetic covermeters. BSI, London
6. British Standards Institution (2004) BS EN 12504: testing concrete: Part 4: determination of ultrasonic pulse velocity. BSI, London
7. British Standards Institution (2012) BS EN 12504: testing concrete in structures: Part 2: non-destructive testing—determination of rebound number. BSI, London
8. Buildings Department (2012) Code of practice for the mandatory building inspection scheme and the mandatory window inspection scheme 2012. Buildings Department, Hong Kong, p 101
9. Bungey JH, Millard SG, Grantham MG (2006) Testing of concrete in structures, 4th edn. Taylor and Francis, Oxon, p 339
10. Chung HW (1994) Assessment of damages in reinforced concrete structures. Concr Int 16 (3):55–59
11. Hong Kong Concrete Institute (2009) Test method for detection of building surface defect by infrared thermography. Hong Kong Concrete Institute, Hong Kong, p 9
12. Hong Kong Concrete Institute (2009) Test method for determination of concrete cover and distribution of steel rebar by surface penetration radar. Hong Kong Concrete Institute, Hong Kong, p 17
13. Hong Kong Housing Authority (2012) HKHA MTS (2012/2014): maintenance and building materials specification: Part D. Hong Kong Housing Authority, Hong Kong
14. Maierhofer C, Reinhardt H-W, Dobmann G (2010) Non-destructive evaluation of reinforced concrete structures. CRC Press, Boca Raton

Upcoming Role of Condition Monitoring in Risk-Based Asset Management for the Power Sector

R.P.Y. Mehairjan, Q. Zhuang, D. Djairam and J.J. Smit

Abstract The electrical power sector is stimulated to evolve under the pressures of the energy transition, the deregulation of electricity markets and the introduction of intelligent grids. In general, engineers believe that technologies such as monitoring, control and diagnostic devices, can realize this evolvement smoothly. Unfortunately, the contributions of these emerging technologies to business strategies remain difficult to quantify in straightforward metrics. Consequently, decisions to invest on these technologies are still taken in an ad hoc manner. This is far from the risk-based approach commonly recommended for asset management (AM). The paper introduces risk-based management as a guiding principle for maintenance management. Then, the triple-level AM model (strategic, tactical and operational) as the foundation to define risk-based AM is described. Afterwards, two categories of risks, one triggered by technical stimuli and the other by non-technical stimuli are introduced. It is shown that the main challenge of managing risks with technical stimuli is to have the ability to understand the technical cause of failures, which is located at the operational level within the triple-level AM model. One method to quantitatively understand the technical cause of failures is by means of condition diagnostic and monitoring technologies. Therefore, the aim of this paper is to clarify the potential contribution of condition diagnostic and monitoring technologies to risk-based decision making for the power sector. This paper shows that, in practice, the implementation of condition diagnostic and monitoring technologies is mainly driven by purely technical asset based considerations without evaluating the contribution to, for instance, risks. This paper provides a list of aspects in which condition diagnostic and monitoring may contribute to risk evaluation with technical stimuli. The listed aspects (which are: (1) asset specific condition data, (2) timely condition data and (3) predictive condition data) can be regarded as input for the probability of failure and as influencing input for the consequence of failure,

R.P.Y. Mehairjan (✉)
Stedin, Distribution Network Operator, Rotterdam, The Netherlands
e-mail: r.p.y.mehairjan@tudelft.nl; ravish.mehairjan@stedin.net

Q. Zhuang · D. Djairam · J.J. Smit
Delft University of Technology, High Voltage Technology and Asset Management, Delft, The Netherlands

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_75

hence benefiting quantitative risk studies and AM activities (such as condition assessment/maintenance or replacement). Finally, these benefits can be evaluated afterwards in a risk-based AM planning stage, so that asset managers can justify investments on necessary technical improvements of condition monitoring systems.

Keywords Asset management · Maintenance · Risk management · Condition monitoring · Electricity networks

1 Introduction

The restructuring and deregulation of the electricity industry has brought about a complete change and presented immense challenges to the electricity distribution network operators (DNO's), regarding their asset and financial portfolios. To meet these challenges, asset management (AM) needs to evolve. In general, DNO's are no longer able to make decisions which are merely technically justified. Examples of decisions that only rely on technical justified reasons could be e.g. "expand the network up to its technical constraints", or "enhance the reliability and redundancy with all available budgets". As a result, DNO's are exposed to two categories of risks, which are:

1. Risks with technical triggers that have economical and societal impacts: *these risks are related to assets*
2. Risks with economical and societal triggers that have technical impacts: *these risks are related to stakeholders*

In Table 1, the above mentioned risks are listed.

Fortunately, it is expected that the evolvement of AM will provide DNO's with capabilities to deal with these risks. As an initial stage of such evolvement, it is seen that DNO's are improving themselves in both business and technology related areas. Firstly, at business level, operation in an electricity market suggests that risk management is one of the key processes for AM decision making for DNO's [1]. Correspondingly, risk management departments and business processes have been established within many DNO's for the purpose of identifying, quantifying, classifying and prioritizing possible risks. Lastly, in the technology area, a revolution is taking place in maintenance strategies, generally speaking, and especially in condition based management methods. The latter is being introduced as a countermeasure to, especially address, the "risks with technical triggers" as can be seen in Table 1.

Preferably, the two above mentioned developments should come about in a related framework, which is commonly regarded as risk-based management (RBM) [2]. However, in the power delivery sector, the situation is that the roadmap to establish this framework is unknown. As a result, it is often encountered that the success of technological developments requires clear links with business level

Table 1 Two categories of risks to which power delivery companies are exposed

Risks with technical triggers that have economical and societal impacts: *asset related risks*	Risks with economical and societal triggers that have technical impacts: *stakeholder related risks*
• Reliability needs to be maintained for the long-term continuity of the DNO	• Investors and creditors expect profitability of the DNO
• The age of components within the network is approaching their design lifetime	• Due to large-scale retirement of employees born in the post-World War II baby boom, a vast loss of expertise is taking place
• New components (e.g. power electronics) are widely installed, but their influence on the existing network is insufficiently understood	• Consumers and the regulator are attempting to control the tariffs of DNO's
• Decentralized generators (e.g. wind turbines) and appliances (e.g. electric vehicles) introduce different load profiles, which require a network of higher capacity or smart used of the existing network to carry them (e.g. dynamic loading)	• Concerns on safety, environment and other public values add to the costs of expansion, reinforcement, maintenance and failure of the network

processes. Therefore, in this article, we aim to clarify the potential contribution of condition based technologies (such as condition monitoring, diagnostics, etc.) to risk-based management in DNO's.

The article is organized in the following way: In Sect. 2, firstly, risk-based management and the role of maintenance are described. Lastly, the triple-level asset management model is described in which the position of risk management is given. In Sect. 3, the role of diagnostics and condition monitoring in risk-based management is described. Finally, as a result, several AM activities in a risk-based regime which can benefit from the application of condition monitoring systems are described to justify investments on condition monitoring technologies. The article comes to a close in Sect. 4 with a number of conclusions.

2 Risk Management in Asset Management—Focus on Maintenance

In today's DNO's, risk-based management (RBM) is seen as a guiding principle within AM strategies [3]. The focus of this section will be on maintenance management as a subsection of AM, and its roadmap towards the link with the risk-based regime.

2.1 Risk Management as Guiding Principle for Maintenance Management

In contrast to what we have seen in the past years, where maintenance management was commonly translated in financial values such as lifecycle costs and total cost of ownership, currently, a trend in the coming years is seen where the value of maintenance will increasingly be quantified in terms of risks. In order to understand this, maintenance organisations will have to evolve towards a certain level of maturity regarding their maintenance regime [4]. From our point of view, the evolvement of maintenance strategies contains the following stages as shown in Fig. 1.

Figure 1 is briefly explained here:

- *Corrective Maintenance* Corrective maintenance is essentially leaving all assets running till failure, and then replacing them. During the time corrective maintenance is being scheduled and performed (usually referred to "break-in", because they "break-in" to the schedule prepared) the asset is inactive [5]. As a general rule [5], a breakdown is often ten times more expensive compared to the situation that the failure can be identified and corrected (or prevented) in a planned and scheduled manner. Until now, the majority of components in distribution networks remains correctively maintained. However, with the adoption of AM, utilities are becoming aware of the changing requirements for maintenance.
- *Preventive Maintenance* The primary upgrade from corrective maintenance to preventive maintenance is by means of maintenance plans and schedules (usually, preventive based on time (Time Based Maintenance), see Fig. 1).

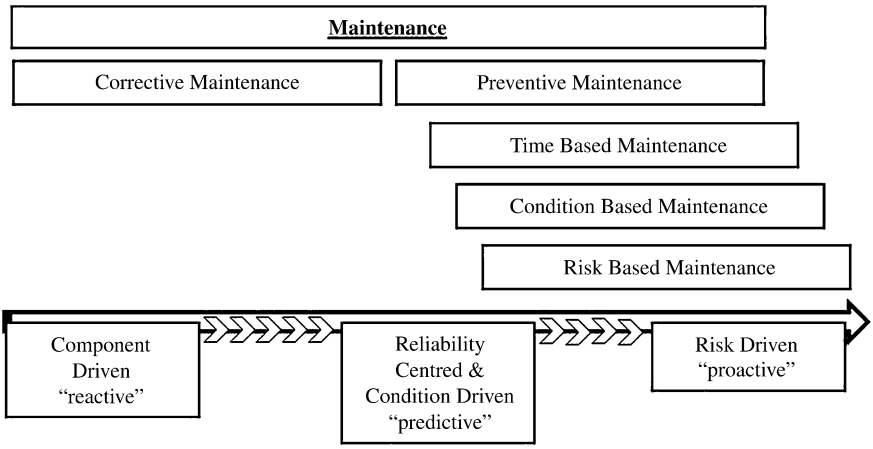


Fig. 1 Evolvement of maintenance management in the last decades from reactive through predictive to finally proactive strategies

Broadly speaking, preventive maintenance plans describe the methods of inspections and maintenance tasks which can efficiently improve the reliability of physical assets. A shift from corrective to preventive maintenance will, inevitably, require some initial investment; however, it will eventually result in moderation of the total volume of planned work and will allow for control of maintenance hours and workload. When preventive maintenance is applied on an asset item it is called a preventive task. Subsequently, the timeline of the preventive task in an asset population is called the preventive schedule. This is why preventive scheduling will, eventually, result in arranging maintenance resources in advance, which, in turn, will considerably accelerate maintenance delivery and reduce operational costs (note, however, that an initial, increased, investment in the transition period is possible, but will decrease once in a controlled period).

- *Time Based Maintenance* as briefly mentioned earlier, traditionally, preventive maintenance is scheduled with predetermined interval, hence the name Time Based Maintenance. The time intervals are decided according to asset type and fixed for the whole lifecycle (usually with reference to manufacturer instructions and updated with historic operational and failure behaviour).
- *Condition Based Maintenance* Basically, condition-based maintenance differs from time-based maintenance in the sense that a shift is made in scheduling methods, namely, from an "intermediately" predictive method to a "fully" predictive method. Being predictive refers to estimating the probability of failures on assets. With condition-based maintenance, an early indication of an impending failure (by applying condition monitoring, diagnostics or inspection methods) can be detected and the consequences of an unexpected failure can be avoided.
- *Risk Based Maintenance* The state of the art maintenance regime is the risk-based version, which is guided by the principles of risk management. A risk is composed of a stimulus (i.e. the root cause) and its consequences. The risk-based approach refers to the quantitative assessments of (1) the probability of stimulus (event), and (2) on business values (Key Performance Indicators, KPI) evaluated consequences. In the planning, the stimuli are the failure modes for risk-based maintenance, which brings the term failure mode and effect analysis (FMEA). In scheduling of risk-based maintenance, the potential failures on asset items are the stimuli. The probabilities of these stimuli are highly recommended to be derived from condition diagnosis (hence, the importance of the upcoming role of condition monitoring in a risk-based management regime). However, in practice, FMEA is mainly based on failure statistics if not expert judgements. The consequences of failure modes and potential failures are, if at all possible, measured with a number of key performance indicators (KPI's), such as customer minute loss, financial loss, safety etc. These KPI's connect the operational-level maintenance tasks with high-level corporate business values. In practice, this link of consequences and failure modes through a certain KPI framework is not

yet straightforward, however this is beyond the scope of this article. Decisions on preventive maintenance plans or schedules are based on the risk register of failure modes or potential failures. Risk register is a process to rank the expected value of risks, while the expected value is the multiplying product of probability and consequence.

From the above review, it can be learned that: the risk-based approach for maintenance is based on five domains of knowledge, which are introduced in different stages of developments. These are:

1. Knowledge of the possibility of failure occurrence (failure modes in preventive maintenance)
2. Knowledge of the measures to prevent a possible failure occurrence (preventive maintenance plan)
3. Knowledge of the approach to predict the probability of a possible failure occurrence (failure statistics or condition diagnosis)
4. Knowledge of the consequences of a possible failure occurrence as well as a KPI system to benchmark this numerically (failure effect in risk-based maintenance)
5. Knowledge of the risk level of a possible failure occurrence (risk assessment methodology, such as risk register)

In the risk-based approach for maintenance, with the five domains of knowledge, two different types of risks can be distinguished of which the first one is with technical stimuli, especially asset failures. This is very familiar to maintenance management. The second type of risk is with financial and societal stimuli e.g. resistance of the public to new substations. The latter risk category initially started to be considered when maintenance management is extended to an organisation-wide asset management approach [4]. In Sect. 2.2 we introduce how these two categories of risk (risk with technical trigger and non-technical risks) are handled in AM.

2.2 *Risk Management Regime in the Triple-Level Asset Management Model*

Asset management is widely accepted and frequently implemented in a triple-level regime. The levels are named *strategic*, *tactic* and *operational* level from the management side to the technical side. Generally speaking, higher levels are concerning wider ranges of assets as a whole, in a longer frame, regarding larger amount of financial investment and consequences. See [3] for a definition for AM for utility companies. The technical and non-technical risk categories can be described for each level (triple-level) of the AM system.

I. The technical risks triggered by failures. These risks are the traditional target of investigation in maintenance management. Additionally, these risks can be

studied scientifically and quantified with probabilities and consequences. Consequently, this allows the classic way to implement risk management and optimize by means of probabilistic data analysis and assess the condition of assets with appropriate technologies (such as condition monitoring tools). In the following, we discuss each level of AM from a technically triggered risk viewpoint.

- At the *operational level*, the hazards of asset failures are investigated, diagnosed and prevented, as the stimuli of "risks".
  - Condition diagnoses are performed to detect failure hazards.
  - The timetable to coordinate preventive maintenance with operation, inventory, human resource, safety measures and other civil works is called maintenance schedule.
- At the *tactical level*, the "risks of asset system failures" are investigated. Accordingly, replacements are scheduled and maintenance plans are decided.
  - The failures of asset systems, rather than assets, are investigated as stimuli of "risks". The term "failure mode" refers to the sequence of aging factor, asset deterioration, asset failure and asset system failure.
  - The consequences of failures are evaluated in several "business values". A business value reflects a KPI of asset portfolio which can be analysed quantitatively and financially.
  - Control of these "technically stimulated risks" is realized through proposal of replacements and decisions on preventive maintenance plans.
  - A preventive maintenance plan specifies
    - which maintenance strategy should be applied on which specific asset, and
    - the diagnosis procedure to be applied on an asset, if it is maintained on-condition.
    - how diagnostic outputs should decide the maintenance schedule.
  - Replacements of long-living assets (typically primary-side high-voltage components) are decided based on a fixed schedule rather than decided risk-based.
- At the *strategic level*, the full spectrum of risks should be managed and controlled.
  - Update the KPI system and review the financially summed risks of asset portfolio.

It is important to stress that from our experience, the main challenge of managing risks with technical stimuli is trying to understand their technical cause (trigger) of failure and the proper diagnosis (predictive) method to understand the condition of assets. This will have to be brought back to a risk-based approach, which will ultimately help to improve maintenance solutions. The left triangle in

Fig. 2 represents the management of risks with technical stimuli. It has a larger area at the operational level (shown as O1), which indicates that the main effort will be made to understand the technical causes of failures. The future trend will more and more require contribution of diagnoses methods, condition monitoring technologies and statistical life data analysis to risk-based maintenance in this area (O1). In this contribution, we will further elaborate on this area within risk-based maintenance and what the role of condition monitoring technologies can be in risk management.

II. The non-technical risk (i.e. societal aspects, such as the development of sustainable energy and the public resistance to power installations). These risks are not considered in the relatively technical maintenance management, but considered by strategy specialists and policy analysers. Additionally, these risks are normally for long-term, so their probabilities and consequences are difficult to predict (e.g. the case that Germany abandoned nuclear power). Consequently, asset managers can only contribute to control these risks through providing innovative technical design of assets/asset systems. In the following, we discuss each level of AM from a non-technically triggered risk viewpoint.

- At the *strategic level*,
  - Analyse future networks, determine risks with commercial or societal stimuli.

    The solutions to these risks are frequently not optimised in standard risk management system such as risk register. Since they are long-term and difficult to quantify (i.e. unlikely to be included in a KPI

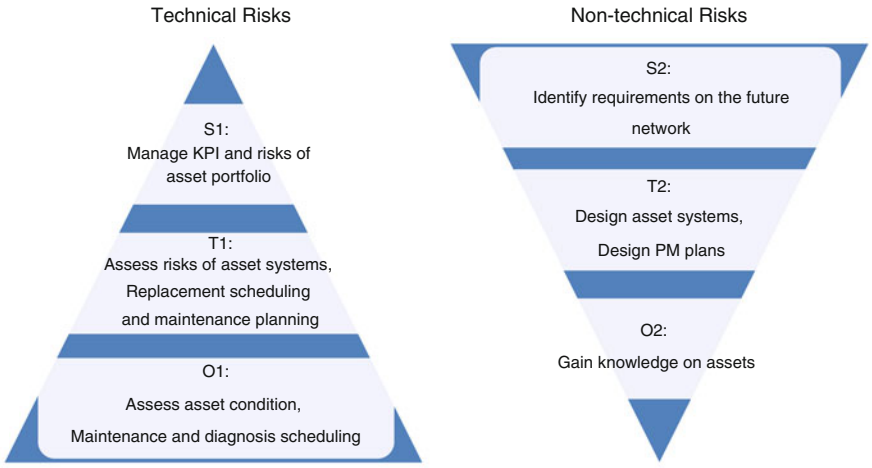


Fig. 2 Represents two categories of risks. The left triangle summarizes the technically triggered risks for the three levels of AM (*strategic, tactical and operational*). The emphasis of this contribution is on area O1. The right triangle summarizes the non-technically triggered risks for the same three levels of AM

system), the asset portfolio should simply be redundant, robust or flexible enough to survive in each scenario.
Such robustness or flexibility can be interpreted technically as hard requirement on asset systems. These requirements are called strategic requirements.

- At the *tactical level*, an equally important task is to design new/replaced/refurbished asset systems, as well as their preventive maintenance plans, so that they can cope with strategic requirements.
- The *operational level* should investigate ways to operate and maintain new components and environments introduced by strategic requirements.

The right triangle in Fig. 2 represents the management of risks with non-technical stimuli. It has a larger area at the strategic level (S2), because the diversity (different specialities) and long term characteristic of these types of risks require a wider human resource (knowledge of overall system) to study.

As mentioned earlier, the technically triggered risk associated with assessing the asset condition will be further elaborated in this contribution. International trends in future maintenance regimes indicate two main developments [6].

Firstly, there is consensus among asset managers that a risk-based approach for maintenance will form the guiding principle in the future. Secondly, the developments in the area of sensor technology and data analysis are rapidly evolving. It is expected in future, that decisions are made based on these facts and figures coming from more asset specific condition assessments rather than on average (degradation) ageing curves. By applying condition monitoring technologies, asset specific risks can be assessed and "moving" risks (the condition of assets vary with time) can be controlled. Yet, in practice, the implementation of wide scale condition monitoring systems are carefully rolled out and usually still as innovation (pilot) projects. Typically, the reason behind this is because it is often unclear on strategic level what the added value of condition monitoring technologies might be. In Sect. 3, we describe the role of condition monitoring in a risk-based management era, and elaborate on the technically triggered risk area O1 (as shown in Fig. 2).

3 Role of Condition Monitoring in Risk-Based Management

In order to incorporate more predictive technologies into risk-based maintenance as a part of an AM strategy, it is vital to demonstrate the value of condition monitoring. To introduce condition monitoring requires sufficient financial investments (stemming mainly from the purchase of condition monitoring systems, employing and training monitoring personnel). To evaluate the added value of implementing condition monitoring, we need to know in which AM activities and at which AM

level (e.g. effective maintenance activities based on monitored condition and on operational level) condition monitoring can improve performance or control risks. Therefore, in this section these activities are discussed.

3.1 General Contribution of Condition Monitoring

Condition diagnostics and monitoring techniques have been applied in the past and gained the interest, especially, at management level [1]. Despite this interest, we see in practice that condition monitoring is occasionally introduced, usually as innovation/pilot project or other reasons such as stakeholder satisfaction after major failures in critical areas. In general, condition monitoring in the power industry has been applied as a method to gather information for the following reasons [1]:

- To manage life expenditures and to ensure that equipment ratings are not exceeded, by monitoring loads and stresses on equipment
- To detect and locate defects or failures. Also, to monitor symptoms of deterioration. This information can be used for the purpose of just on time warnings and as data for condition assessment for guiding maintenance and replacement activities, hence supporting AM decisions, especially, on operation level.

The interpretation of data coming from condition monitoring systems, the reliability miss-match of diagnostic systems with the equipment being monitored and the volume of data (big data challenge) damped the application of condition monitoring systems. Another important issue is the timeliness with which the acquired condition data can be provided and the relationship with the time to failure of this specific asset [7]. Nowadays, most of these challenges remain and form an obstacle for large scale applications of condition monitoring, especially in combination with the costs for setting up a condition monitoring programme. Due this, it is often unclear to asset managers what the added value of condition monitoring systems are, especially in terms of potential benefit to risk-based management.

However, the above mentioned obstacles can be avoided in the risk-based AM regime. In the next section the role of condition monitoring within the scope of technically triggered risk stimuli on an operational level, represented in Fig. 2 in the left triangle, is explained.

3.2 Role of Condition Monitoring Within Risk-Based Management

In Fig. 2 we explain that the focus in this contribution is on the left triangle (risks with a technical trigger) and especially on the O1 (operational) area where the assessment of asset condition forms an important part. In order to fulfil the tasks for

Table 2 Detailed listing of the role of condition monitoring to technically triggered risks

Contribution of condition monitoring to the assessment of risks with technical stimuli
Asset specific:
Contribute to a specific asset service condition and remaining life assessment
Contribute to sub-systems (families of population) of assets long time condition behaviour assessment
Contribute to gain knowledge of measured condition in the whole lifecycle of assets (e.g. assessing the changing risk of failure of critical components based on whole lifecycle condition data)
Just in time reaction:
Contribute as warning as an input for alarms for timely made decisions for preventing failures
Contribute to environmental hazards prevention such as warnings for harmful substance release (this can additionally be used in the non-technical risk stimuli)
Predictive performance contribution:
Contribute to obtain predictive information about the degradation of assets operating in the network. This can be useful for identifying critical service condition for equipment.
Contribute to study the impact of environmental influences on the condition of assets

the assessment of asset condition, it is required to have insight (information) of the following aspects [7]:

- Technical knowledge of the component
- Functional description of the component
- Stresses which are imposed by loads or environments
- How these stresses deteriorate the components

In quantitative risk studies, measurable data is required to determine the equipment condition, probability of failure and associated risk(s). In order to calculate the probability of failure requires statistical failure analysis. However, to meet the requirements stated above in order to fulfil the tasks for the assessment of asset condition requires the application of at least some form of diagnostics or condition monitoring.

In Table 2, we list the aspects in which condition monitoring contributes to the assessment of risks with technical stimuli, hence contributing to the operational level of the left triangle shown in Fig. 2.

In general, this list can be regarded as input for the probability of stimulus (root cause) and as influencing input for the consequence of failure (impact of a failure). The consequences can be reduced because component deterioration can be remedied before, for example, safety is affected, service is interrupted or significant damage occurs. This is explained as follows:

- Regarding the probability of stimulus (failure mode)
  - Reducing equipment major failure probability
  - Preventing extensive life cycle loading and/or temporary overloading of an equipment

  - Disclosing already deteriorated equipment conditions before they develop into a major failure and cause unplanned outage

- Regarding the consequences of failures
  - Preventing high cost of major and fatal failures equipment repair (incl. replacement)
  - Preventing consequential damage of neighbouring equipment
  - Controlling outages (planned outages)
  - Lowering insurance fee at insurance companies

4 Summary and Discussions

An internationally observed trend within asset management is to adopt risk-based approach as a state-of-the-art and cost effective maintenance regime to control risk profiles. This is widely accepted and applied by utility networks nowadays. This paper aims to find out how condition diagnostic and monitoring technology can contribute to risk-based management in two steps.

The first step is to reveal how specific RBM activities benefit from condition diagnostic and monitoring technology. In order to locate such activities within the RBM framework, we firstly divided the RBM framework in two dimensions: (1) the three different AM levels—*strategic*, *tactical* and *operational*, and (2) the two different categories of risks—*technically triggered* and *non-technically triggered*. After introducing these levels and categories, we have identified that condition diagnostic and monitoring systems will mainly contribute to the operational AM level when technically triggered risks are managed. By applying this, the technical hazards can be quantitatively assessed and maintenance activities, as the countermeasures, can be further optimized.

The second step was to propose how condition diagnostic and monitoring systems can facilitate quantitative risk assessment through proper management on information acquired from them. We provide a list of aspects that contribute to risk evaluations with technical triggers. The listed aspects are: (1) *asset specific condition data*, (2) *timely condition data* and (3) *predictive condition data*. These can be regarded as input for the probability of failure and as influencing input for the consequence of failure, hence benefiting quantitative risk studies and AM activities, such as condition assessment/maintenance or replacement.

As a consequence, when above mentioned two steps are taken into account, asset managers can evaluate the benefits afterwards in a risk-based AM planning stage. Moreover, such evaluations can help to reconsider decisions on necessary technical improvements of condition diagnostic and monitoring systems and to justify future investments into these systems.

References

1. Cigre WG (2006). C1.1: asset management of transmission systems and associated cigre activities. Cigre Technical Brochure vol 309, December 2006
2. Cigre WG (2010).C1.16: Transmission asset risk management. Cigre Technical Brochure, vol 422
3. Rijks E, Southwell P (2010) Asset management strategies for the 21st century. Electra 248 (1):29
4. Mehairjan R et al. (2012) Organisation-wide maintenance & inspection improvement plan: a dutch electricity & gas distribution network operators approach. IET Asset Management Conference. London
5. Wilson A. (1999). Asset maintenance management: a guide to developing strategy and improving performance. ISBN 0-9506-465-3-9. Conference Communication 1999
6. Baarle van D (2013) (in Dutch). Kern van Onderhoud is Kennis, pp 46–54. iMaintain Magazine, the Netherlands
7. Cigre WG (2011). B3.12: Obtaining value from on-line substation condition monitoring. Cigre Technical Brochure, vol 462. June 2011

Enhancing the Management of Hong Kong's Underground Drainage and Sewerage Assets

Stephanus Shou, H.S. Kan, Martin Jones, Craig Roberts and Andrew Tsang

Abstract The Drainage Services Department (DSD) is developing a long-term holistic management system for sewer and drain replacement and rehabilitation. The system will utilize a risk-based approach to prioritize rehabilitation and replacement (R&R) works and optimize preventive maintenance inspections and surveys. The project involves a review of existing asset information and asset management system, and the development of decision support tools for risk classification and works prioritization. Data in-filling techniques will be utilized to deal with data gaps, and deterioration models will be developed to project the latest structural conditions of assets. In addition, advanced inspection and rehabilitation technologies will be evaluated for possible pilot trials and applications. This paper will describe the approach in developing the system, work done to date, and how the system will tie in with the wider asset management initiatives of the organization. It will also discuss how the new system will enhance the operation and maintenance and services of DSD.

Keywords Asset management · Sewer rehabilitation · Data infill · Deterioration modelling · Risk-based planning

S. Shou (✉) · M. Jones · C. Roberts
Black and Veatch, Kansas City, USA
e-mail: ShouWL@BV.com

M. Jones
e-mail: JonesM2@BV.com

C. Roberts
e-mail: RobertsC@BV.com

H.S. Kan · A. Tsang
Drainage Services Department,
The Government of the Hong Kong Special Administrative Region,
Hong Kong, China
e-mail: hskan@dsd.gov.hk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_76

1 Introduction

The underground assets of DSD have expanded rapidly with the continuous development and urbanization of Hong Kong in the past few decades. As of year-end 2011, DSD was responsible for maintaining an underground sewerage network comprising approximately 1,540 km of pipes, 20 km of box culverts, 198 km of rising mains and an underground drainage network of approximately 1,820 km of pipes and 410 km of box culverts. In 2011, there were several incidents of road subsidence caused by the collapse of aged and dilapidated sewers and drains. DSD determined there was a need for enhanced management of the underground sewers and drains, incorporating a systematic and well-organized inspection program, and development of an R&R strategy.

To achieve the above objectives, the main tasks are to review the existing asset management system and to enhance it to allow the prioritization of R&R works using a risk-based approach. The whole process comprises two phases. Phase 1 is to conduct data collection, data review, data estimation and updating, and to develop methodologies for prioritization of R&R works. Phase 2 mainly involves R&R works prioritization, option analysis for prioritized R&R works and formulation of an R&R Management Plan. Phase 2 is scheduled to commence in March 2014 for completion in April 2015.

The risk-based approach to prioritize R&R works will take account of the likelihood and consequence of pipe failure. Adoption of the approach starts from the collection and mining of existing asset data, which include age, material, size, surrounding environment, structural condition etc. The asset data provides the fundamental information for the whole sewer and drain networks which will inform the estimation of failure likelihood and consequence for assessing the overall risk. The data was subsequently reviewed to identify data gaps, and then data infill techniques were applied to fill the data gaps in order to support the risk-based prioritization approach.

DSD has been implementing a five-year to ten-year inspection CCTV survey cycle aimed at covering the whole sewer and drain networks. Despite this, part of the networks cannot be examined due to various site constraints including traffic, high flow conditions and lack of access points. Deterioration models will therefore be used to estimate structural condition and failure likelihood for assets with no or very old observed condition grade.

Before applying the infill techniques and deterioration models to the whole territory, a sub-district in Tuen Mun (TM2) was selected as a pilot area for implementation in order to trial the proposed methodologies for data infill and deterioration modelling. The following sections will provide details of the works done so far and the way forward based on the findings from the review of the TM2 sub-district.

2 The Importance of Good Data to Support the Risk-Based Renewal Planning

To develop optimal risk-based asset management plans (i.e. renewals planning) the availability of good quality asset related data is vital. In particular, for renewals planning purposes, DSD is using a risk-based approach to answer the following questions:

- What is the strategy to effectively manage the asset risk and target the critical assets?
- What are the probability of failure and the consequence of failure of the assets (risk)?
- What are the appropriate interventions (replace, rehabilitate, repair, observe) to manage risk and maintain service levels?
- How much will it cost (capital and operating costs)?
- How much asset replacement, rehabilitation or repair (activity) is required to achieve the service targets, balance risks and the budget?
- How will the proposed interventions and activity affect service levels, e.g. the residual benefit?

To address these questions analyses of good quality asset related data is required. Analysis will include determining probability of asset failure, consequence of asset failure and impacts of different interventions on the asset risk, service levels and performance. The results of this type of analysis will be used to forecast targeted interventions and activity rates, future levels of service and estimate required capital investment and operating costs. Good quality outputs from the data infilling exercise and the deterioration modelling are therefore a critical aspect to the success of the overall R&R risk-based planning approach.

3 Data Infill Approach

3.1 The Current Data Situation

The basic asset attributes include:

- Age
- Structural condition
- Material
- Size
- Invert level

Age and structural condition are crucial deterministic factors in risk-based planning. For these, data gaps are observed and data infilling techniques have to be explored.

The chosen data infilling techniques will be applied to a pilot district in Tuen Mun, known as TM2, as this area has a relatively high availability of asset age information close to 80 %. It will serve as a good representative area to examine various infill techniques, in particular those requiring spatial relationships.

3.2 The Data Infill Methodology

Initially all the available pipe characteristics and performance related data for the whole of Hong Kong (not only the TM2 sub-district) were collected and consolidated into a central geo-database. This allowed the analysis of statistical distribution and spatial relationships in the data, which would inform the data infilling and subsequent deterioration methodology. An example is shown in Fig. 1 which depicts how known age data is spatially distributed across the territory.

Both positive (i.e. relationships that can be used as a factor in the infilling) and negative (e.g. indicates lack of relationship) correlations have been identified. An example of a statistical distribution which can be used to inform data validation or integrity checks is shown in Fig. 2. It shows that an integrity check can be used to flag any Vitrified Clay pipes having over a 600 mm diameter for further investigation. Unfortunately the results indicated limited opportunity to use statistical distribution to infill. Some of the results will be used for validation purposes.

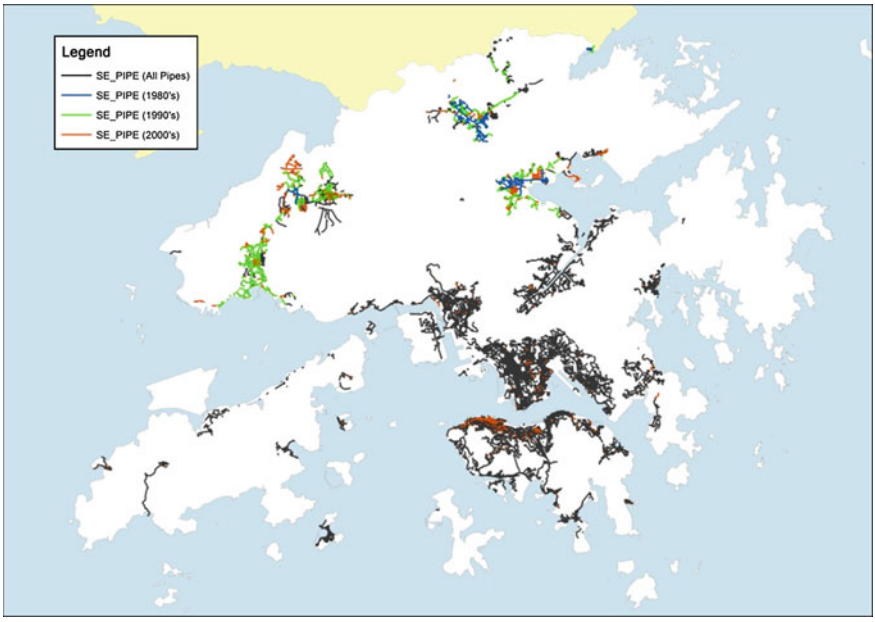


Fig. 1 Availability of age date in whole territory

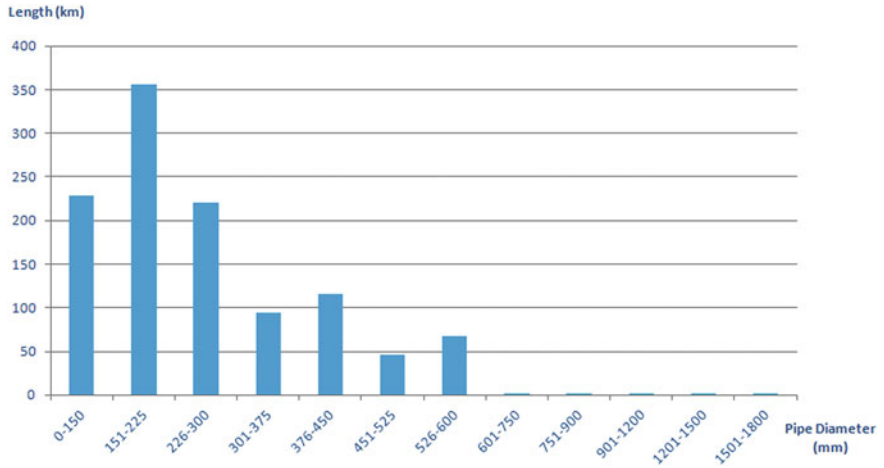


Fig. 2 Size distribution of vitrified clay pipes

Spatial distribution analysis has resulted in a number of infill techniques being established which are introduced below.

The source of ancillary data have also been investigated which would be beneficial to the gap filling exercise. This is mainly focused on infilling age data on account of its relatively low availability and spatial distribution. Some of the ancillary data investigated in the pilot has included the Water Supplies Department's water network, building age from the Lands Department, customer data and Airborne Light Detection and Ranging (LiDAR) Data (for invert level estimation). After testing the correlation of age from the various ancillary sources with known age data, the building age data were found to be the best ancillary at this stage of work. Sourcing better age data is still ongoing and data from the Hong Kong Road Maintenance System may also be reviewed. Figure 3 shows the good distribution of building age across Hong Kong. Invert level estimation using LiDAR has proven to be unreliable resulting in instances of downstream invert levels being higher than known upstream inverts.

Eleven infill techniques have been developed and tested in the pilot phase. The techniques vary widely from textual searching of historic work management data to performing spatial analysis, such as clustering, tracing and nearest neighbour. The majority of the techniques have some forms of spatial analyses applied.

Only limited validation checks and no automatic correction were performed for the pilot as the main focus was to develop the infill methodology. During the main phase work validation checks will be further developed and run prior to and post infill. Appropriate automatic correction rules will also be developed in the main phase.

The eleven infill techniques developed and applied in the pilot are shown in Table 1 above. They were used to infill age, size and invert level for the sewer and drains pipes, box culverts and rising mains. The Reliability, Accuracy and

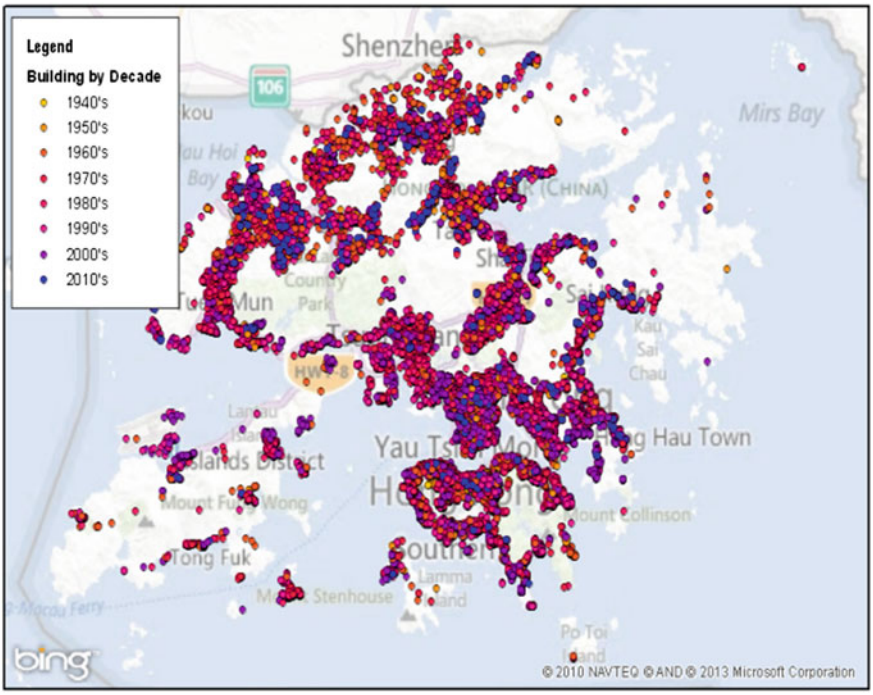


Fig. 3 Building age across Hong Kong

Table 1 Data infill techniques

Technique	Reliability	Accuracy	Confidence grade (CG)
Supplied data	A	2	A2
As-built drawing data	A	1	A2
(Size) work order description field	A	2	A2
(Size) retrieve upstream	B	3	B3
(Size) retrieve downstream (<25 m)	B	3	B3
(Size) retrieve downstream (>25 m)	B	4	B4
(Size) nearest neighbour	C	4	C4
(Age) upstream clustering	B	2	B2
(Age) nearest building age	C	3	C3
(Age) nearest 10 pipes	C	4	C4
(Age) thiesen polygons	C	5	C5
(Invert) known level connected	B	2	B2
(Invert) known level tracing	B	3	B3
(Invert) LiDAR average depth	C	4	C4

Confidence Grade columns refer to the data confidence grades assigned to infill techniques. These grades are based on the UK OFWAT data confidence grades [1]. They will be used to estimate the resultant quality of the data estimated by the infill techniques. This is important as the data are used as a deterministic factor in the subsequent risk-based planning exercise. The Reliability (A–D) assessment ranges from sound textual records, data, investigations, works and analysis through to unconfirmed verbal reports. Accuracy ranges from 1 to 6 and X, and is defined as a likely percentage ∓ level of accuracy (e.g. 2 is ±1–5 %). These data confidence grades will be applied to every attribute, combined and then normalised for each individual asset in the main phase. In addition the confidence grade was used to decide the priority of infill techniques.

As can be seen from Table 1 varying levels of confidence have been assigned through testing the accuracy of the infill technique using statistical population tests against random chance and by applying the authors' expert judgment. For example the technique to infill age by spatial clustering and upstream tracing has a B2 data confidence which is a good grade. This is due to the accuracy test indicated an over 90 % chance the estimate of age will be correct compared to approximately 3 % for random chance.

Currently the building age infill technique, which has a C3 (extrapolation from limited sample of Grade A or B data and an accuracy band of ± 10 %) data confidence is statistically useful but ideally should be improved. This technique attempts to infill the age by spatially analyzing all buildings within 100 metres of the asset. The nearest building which has an age is then used to infill the age. The data confidence and the infill technique used are also associated with the individual asset record in the central database. Figure 4 depicts the building age infill technique.

After running the infill method on the pilot area all of sewer and drain assets with missing age, size and invert level were infilled. Each attribute infilled was also associated with its data confidence grade and the infill technique name. The pilot infill results show that more than half of missing data except the invert level were infilled by methods with reliability Grade B or above. The accuracy of infilled data is satisfactory and the data will be usable in the coming risk assessment. Fine tune of the infilling techniques for invert level to improve the accuracy will be needed in the main phase.

4 Deterioration Modelling for Tuen Mun Pilot Study

One of the key components of asset management system is the ability to predict the infrastructure's future performance. Ana and Bauwens [2] Deterioration modelling is therefore introduced for use to predict the structural conditions of sewer and drain networks to allow better planning of maintenance activities including inspection, rehabilitation and replacement.

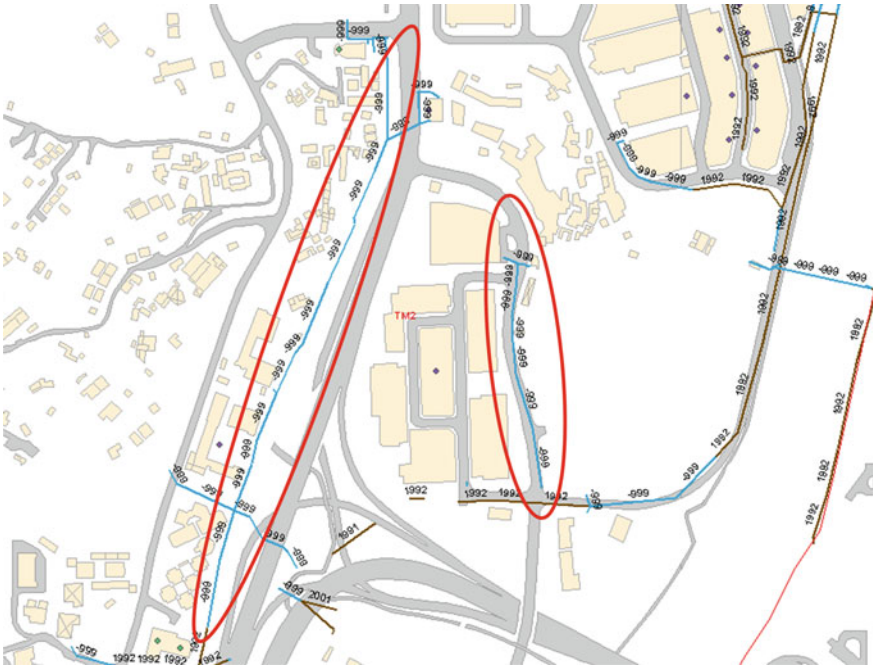


Fig. 4 Building age infill technique

Deterioration models can be used to estimate the condition and likelihood of failure based on available asset characteristics and failure data. There are a number of different methodologies available for forecasting pipe deterioration, and they are dependent on the type and quality of available data. Deterioration models can be developed based on remaining expected life, asset condition grade and service levels, such as collapse rates, blockages and rising main failures.

The approach to the pilot study was to review the available data for developing and applying deterioration models, trial different methodologies and determine the most appropriate methodology to use. The type of model and the accuracy of predictions are very dependent on the availability and quality of data. Table 2 summarizes the typical data required for a deterioration model and the availability of them for Hong Kong.

Asset data of notable levels of availability for developing deterioration models are limited to age, material, diameter and condition grade. The next step was to review different deterioration methodologies to determine their suitability based on the available data, and then develop initial models to trial. Four methodologies reviewed are summarized below.

Weibull Distribution The Weibull distribution is commonly applied in reliability analysis to model distributions of times to events (failures). It is good for

Table 2 Deterioration model data requirements

Data Required	Relevance
Date of construction	Age is an important factor in the deterioration model, although it is not necessarily the dominant factor
Material	Typically different materials deteriorate at different rates
Diameter	Size can be a factor in deterioration rates
Soil type	The type of soil surrounding a pipe can be a significant factor
Pipe bedding	Pipe bedding can be a factor in relation to pipe deterioration
Traffic loading	Traffic loading can be a factor in relation to pipe deterioration
Condition grade	This data if available will be used directly and deterioration modelling need not be applied
Service level	Examples of this include reported collapses, rising main bursts and blockage

network-level pipe understanding and capital budgeting, and relatively simple to apply. However, it requires good age data.

Markov Chain The Markov Chain is a discrete-time stochastic process, where the conditional probability of the future state only depends on the present state. It is often used in deterioration modelling as it can model the transitions from one condition grade to another. It requires repeated survey data to determine the change in condition grade on an individual pipe over an appropriate period of time. However, there seems to be insufficient repeated survey data in Hong Kong to apply this methodology.

Multinomial Logistic Regression This regression methodology is used to model categorical variables like pipe condition grade. It provides the probability that a pipe with given characteristics will fall into a certain category, and can be applied at pipe level.

Probabilistic Neural Networks Neural networks are a form of artificial intelligence or soft computing that attempts to mimic the way people learn. Many different pathways are attempted with those leading to a correct answer in a training set given increased weight. This approach is capable of considering a wide variety of categorical and numerical data and using the data to generate condition rating probabilities. There are extensive data and computational requirements for this methodology.

The review identified the Weibull and Multinomial Logistic Regression methodologies as the most appropriate ones to develop. The in-filled data from the pilot area and available data from the whole of the Hong Kong were collated and analysed prior to the development of the models. It should be noted that at this stage only a limited amount of age data has been in-filled for the whole of Hong Kong, and further work is needed to complete the data infilling. The analysis identified a relatively high rate of deterioration of the sewers and drains in the pilot area. Condition grade data for the rest of Hong Kong shows a more gradual deterioration rate. Figure 5 shows pipe condition by age at inspection for sewers and drains which have both an age and an observed condition grade, with the percentage of pipe in condition grade 1 to 5 by age of the pipes (grade 1 is very good condition

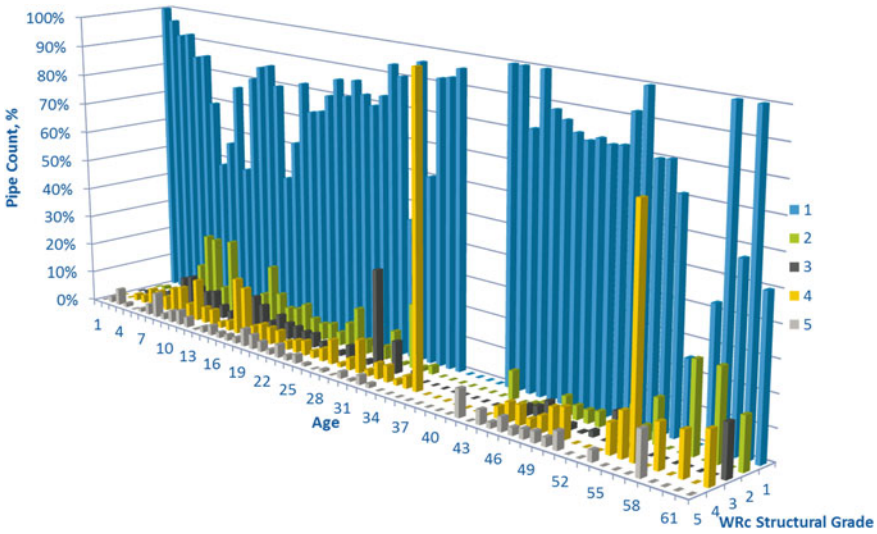


Fig. 5 Condition grade by age of sewers and storm drains

and grade 5 is very poor). It is expected that the percentage in grade 1 would reduce over time, with the percentage in grades 4 and 5 slowly increasing. However, this is not apparent from the chart, and the data do not give a clear picture of deterioration over time.

The Weibull deterioration models were applied to sewers and drains in the pilot area which did not have an observed condition grade in order to predict the probability the pipes were in condition grade 4 or 5. The results were validated by running a random 20 % sample of pipes with known condition grades in the pilot area and comparing the known grades with the outputs from the model.

The results showed some correlation between predicted and actual structural conditions, but further work is needed to improve the model accuracy. The current data set is limited, and this should be improved once the remaining gaps in the age data for the whole of Hong Kong are filled using the in-filling techniques discussed earlier. The different deterioration rates between some districts need some further investigation, as well as the option of developing separate models for different cohorts of pipes with similar deterioration characteristics (Fig. 6).

5 Concluding Remarks

In order to develop a risk-based approach to sewer and drainage R&R prioritization, extensive asset data is required to inform the likelihood and consequence of failure for the risk assessment. Analysis of the Hong Kong asset data identified a number

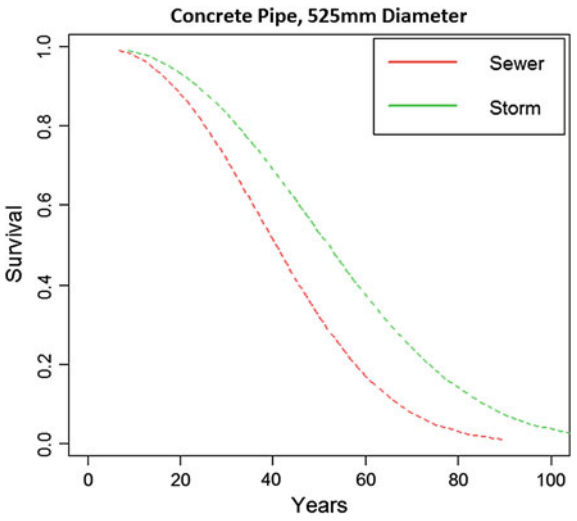


Fig. 6 Concrete pipe deterioration

of gaps, which require the application of data infill techniques and deterioration models to estimate the missing data. Eleven data infill techniques and two deterioration modelling methodologies were trialled in the Tuen Mun pilot study area to assess their suitability.

Data infilling is largely dependent on spatial techniques (e.g. tracing, clustering) using the available data and the sourced ancillary data. Initial estimates of infilled data confidence are relatively low for some of the techniques applied to estimate asset attributes, such as invert level. Further work to improve the techniques where they are contributing significantly to data infilling will be required prior to application to the whole of Hong Kong, particularly for asset age which is an important attribute for deterioration modelling.

The Weibull methodology was identified as the most appropriate deterioration methodology, but the available data does not give a clear picture of deterioration over time, and further work is needed to develop the deterioration models using an improved data set.

The next stage is to apply the in-filling and deterioration methodologies to the whole of Hong Kong, and then use the in-filled asset data to inform failure likelihood and consequence for the assets, and to undertake a risk analysis. The results from the risk analysis will provide each pipeline a risk score which will be used to prioritize asset inspection, replacement and rehabilitation, and to develop a long-term R&R implementation programme.

References

1. Office of Water Services (2011) Reporting Requirements Version 1.1. http://www.ofwat.gov.uk/regulating/junereturn/reportingreq/. Accessed June 2011
2. Ana EV, Bauwens W (2010). Modeling the structural deterioration of urban drainage pipes: the state-of-the art- in statistical methods. Urban Water J 7(1): 47–59

Implementation of Computerized Maintenance Management System in Upgraded Pillar Point Sewage Treatment Works

Henry K.M. Chau, Ricky C.L. Li, Tim S.T. Lee, Bill S.M. Cheung and Teck Suan Loy

Abstract In the past, monitoring and scheduling the operation and maintenance activities of physical assets in Sewage Treatment Works (STWs) of Drainage Services Department (DSD) follow a traditional risk based approach with due consideration to the financial as well as the state of the assets. For newly Upgraded Pillar Point Sewage Treatment Works (PPSTW), Recursive Auto-Regression (RAR) modelling [1] technique is adopted to automatically predict specific equipment's remaining useful life (RUL) and compare the lead time of components' delivery and process time of overhaul sub-contracting so as to establish an optimum preventive maintenance schedule with delivery, resources and cost optimization. Prediction accuracy of the developed RAR model is verified by numerical simulation with inputs to CMMS condition monitoring engine. A pilot study on the integration of CMMS with the SCADA system has been implemented on the outfall screw pump shaft bearings for experimental validation of the RUL model and investigating the feasibility of its application.

Keywords CMMS · Recursive auto-regression · Condition-based monitoring · Remaining useful life

H.K.M. Chau · R.C.L. Li
Drainage Services Department, Government of the Hong Kong Special Administrative Region, Hong Kong, China
e-mail: henrykmchau@dsd.gov.hk

R.C.L. Li
e-mail: rickyli@dsd.gov.hk

T.S.T. Lee · B.S.M. Cheung (✉)
AECOM Asia Company Limited, Hong Kong, China
e-mail: sre@dc200803.com

T.S.T. Lee
e-mail: tim.lee@aecom.com

T.S. Loy
ATAL—Degrémont-China State Joint Venture, Hong Kong, China
e-mail: teck.suan.loy@degremont.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_77

1 Introduction

Traditional risk based approach of a labour intensive data logging, tracking and analysis exercise by experienced staff is currently adopted for monitoring and scheduling the operation and maintenance activities of physical assets in sewage treatment facilities of DSD. With the development of information technology, there is an increasing trend that a more systematic management of such assets, i.e. asset management is desirable to achieve an efficient and effective operation of the STWs. For the "Design, Build and Operate Pillar Point Sewage Treatment Works" project, DSD has implemented seamlessly asset management strategy and philosophy with integration of SCADA system and CMMS system for optimization of life cycle cost, finance, system reliability, work force, inventory, resources, etc.

2 Asset Management

As defined (PAS 55, 2004), Asset Management is "systematic and coordinated activities and practices through which an organization optimally manages its assets, and their associated performance, risks and expenditure over their lifecycle for the purpose of achieving its organizational strategic plan." [2] which is of growing importance to contribute better services with enhanced performance and efficiency. Asset management combines management, finance, economic, engineering, risk and reliability, and other factors that are to be applied to physical assets in providing required level of services throughout an asset's service. It is a systematic process of managing the asset's entire life cycle in most cost-effective and optimized manner, including its design, construction, commissioning, operation, maintenance, repair, modification, replacement and ultimate disposal [3]. Normally asset would remain in working for a substantial period of time when best practices in proper operation and maintenance are adopted. Subject to the assets usage, nature, quality and its operation environment, it is of vital importance that well-timed maintenance and routine inspection for achieving an optimum service life of an asset with the aims of expert knowledge before catastrophic failure occurs [4].

Asset management philosophy and strategy for Upgraded PPSTW are developed based on the following objectives:

- Failure prediction and prevention
- Equipment redundancy and availability
- Condition-based maintenance with continuous monitoring and assessment
- Asset performance benchmarking
- Optimization of capital expenses (Capex) and operation expenses (Opex).

Capex is the cost of developing or providing non-consumable parts for the product or system while Opex is an ongoing cost for running a product, business or system [5]. In other words, Capex and Opex constitute all costs that are allocated to

an asset throughout its entire life cycle. The asset management strategy integrates Capex and Opex optimization to deliver maximum cost saving by extending the useful life of the equipment with predictive and preventive maintenance approach while corrective maintenance is minimized.

All in all, integrated control, information and database system, i.e. Asset Management Package, CMMS Package and SCADA Package, plays an important role for the implementation of a sound meaningful asset management system. In modern sewage treatment works, Supervisory Control and Data Acquisition (SCADA) system, as its name implies, is the key element for process automation and data gathering by a computerized and centralized industrial control system that monitors and facilitates control of the whole works. Majority of the control actions are performed by remote terminal unit (RTU) or programmable logic controller (PLC) which control functions are normally limited to supervisory level intervention and override, that is, RTU/PLC processes the feedback control loop and the SCADA supervises overall performance of the loops. Another function of SCADA system is data acquisition which initiates inputs at field side such as meter readings, equipment status, etc. that are connected to SCADA. Data is then transferred and visualized at the plant control room via Human-Machine Interface (HMI) for subsequent plant operation and decision making [6]. However, the SCADA system alone does not possess any planning, scheduling, monitoring, etc. of asset management and maintenance functionality.

Proper maintenance and appropriate renewal of parts, equipment and plant are considered as one of the best practices to extend the service life of an asset in order to arrive at the lowest lifelong cost. For implementing Design, Build and Operate contract in Upgraded PPSTW, asset renewal cost forms an important element of the overall cost and must be properly accounted for so that the planned maintenance and renewal can be undertaken. CMMS is typically adopted and widely used to keep the track record of all entities within the Works which is a proprietary software package that maintains database of an organization's maintenance operations and facilitates the management of plant's system and equipment, daily operation work, correction and preventive maintenance works. The CMMS, which is capable in creating, linking, triggering and maintaining Corrective Maintenance (CM) and Preventive Maintenance (PM) tasks, is therefore implemented with Maintenance, Inventory and Procurement Modules to facilitate the management of assets for Upgraded PPSTW.

The architecture of Upgraded PPSTW's asset management system is presented in Fig. 1. It is a third-generation networked architecture of four layers, namely, field bus, control network, plant network and business network. Field bus refers to all available field sensors, instrument readings, equipment status, running information etc. that connected to the PLC of each treatment process unit with various types of contacts (dry contacts, hardwires, etc.) and buses (Modbus, Profibus, etc.). The PLCs are then configured to form a complete control network of the Works in accordance with developed system control philosophy. The SCADA system connects the control network for overall plant control, monitoring and data logging and visualizes the plant operation with HMI. The business network consists of CMMS

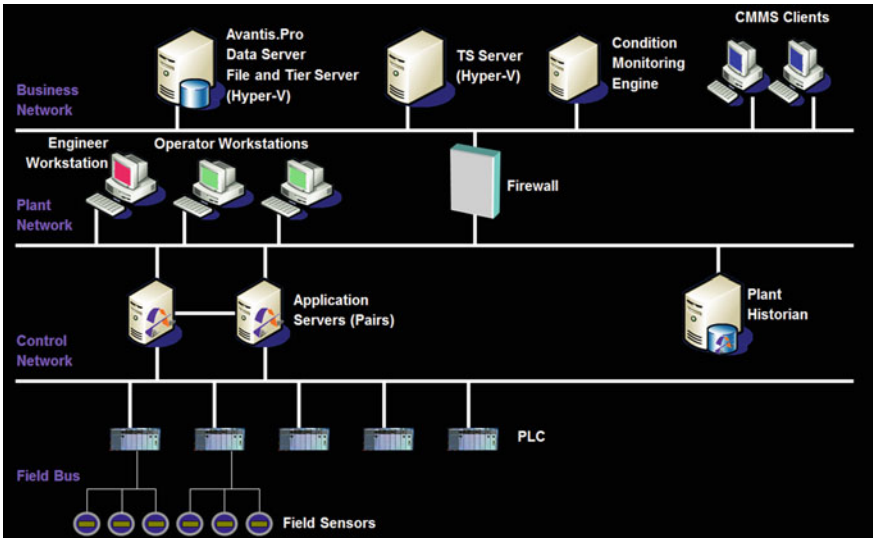


Fig. 1 Asset management system network architecture

server and condition monitoring engine to form the business network that is attached to the plant network and integrates with the SCADA system. It utilizes standard communication protocols and security techniques when implementing the integration of SCADA, CMMS and Condition Monitoring Engine. To reduce the potential vulnerability of the Asset Management System from remote attack, accessibility over internet is denied. To restrict un-authorized access or tackle an attack, hardware-based firewall is also installed which controls incoming and outgoing traffic between CMMS and SCADA networks and secures the network with trust.

Condition based monitoring engine integrates real time plant information and assesses equipment status accurately and timely. It relies on rule-based condition monitoring engine for creating work orders automatically that minimize human and capital resource. The CMMS database file server, tier server and remote desktop services server (Hyper-V) utilize Microsoft virtualization environment to achieve central administration tasks which host CMMS database, middle tier and transaction services components via integrated CMMS operation environment with computer. Data Flow of Condition Based Monitoring System is illustrated in Fig. 2.

3 Computerized Maintenance Management System

The core of the CMMS of Upgraded PPSTW consists of Maintenance, Inventory and Procurement module which aims to facilitate the management of plant's system and equipment, daily operation work, correction and preventive maintenance works

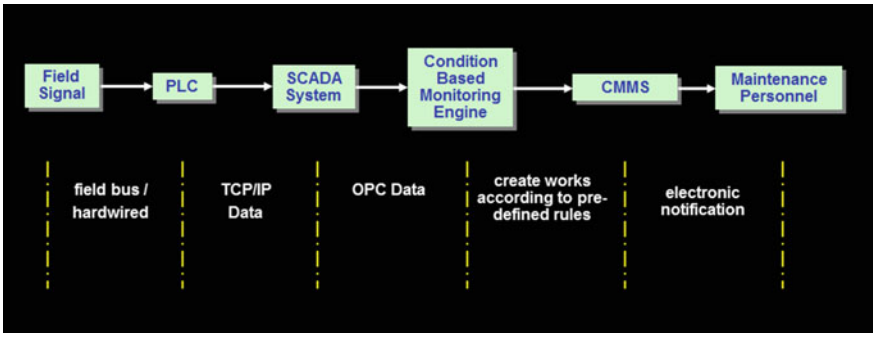


Fig. 2 Data flow of condition based monitoring system

and generation of various analysis and management reports. The CMMS creates, links, triggers and maintains CM and PM activities and also keeps track of status of Work Order.

The CMMS features the following Maintenance Management functions:

- Entity Management (Equipment Costs and History tracking)
- Work Planning (Work Order creation, planning, printing, tracking and closing)
- PM Jobs (Creating PM jobs, assigning frequency of PM to Equipment, triggering and executing PM Work Orders)
- Maintenance, Repairs, and Operations (MRO) Inventory Management

Entity management is incorporated in the CMMS to maximize the flexibility of equipment database, which supports the management of both maintainable and organizational entities, as well as tracking capital projects and related tasks and expenditures. Maintainable assets include anything that needs to be repaired such as a piece of equipment whilst organizational entities refer to any arrangement of the system that is used to collect cost, statistical, budget or backlog information such as a cost centre. For Upgraded PPSTW, the CMMS is capable to keep track of all the spare parts that are being used in those maintainable entities and organizational entities created within the CMMS.

The Work Management of CMMS is purposed to ensure that maintenance personnel can manage and plan incoming work requests as well as automatically generated work from preventive maintenance programs. Work Planning is essential to ensure that labour, materials, tools, drawings and subcontractor requirements, as well as safety information, can be identified on work orders to support proactive maintenance activities. For Upgraded PPSTW, the CMMS enables operator staff to use the work request as a simple electronic tool to communicate a need for service to the maintenance department by creating Work Orders from pre-planned Work Order Templates. Work Orders are created in several ways, namely, Simple Work Orders that contain only one task for a single entity while Multi-Task Work Orders are adopted when multiple entities are affected or multiple task are required. The CMMS identifies the entity to be repaired, overhauled, replaced or maintained on

the Work Request and Work Order with condition monitoring features incorporated. In addition, to ensure the safety of maintenance personnel, the CMMS generates and provides safety instruction of any potential hazardous or dangerous condition associated with a job, the entity on which the works is to be performed, or the materials or tools involved with the work. The CMMS also provides with the material transactions functionality which consists of issues and returns from the warehouse for inventoried items and procurement for non-inventoried material. When closing a work order, the CMMS records the entry of allocated costs with descriptions for future historical analysis.

The CMMS also features with Planned Preventive Maintenance (PM) function which is based on user defined activities to be performed with essential resources input such as labour requirements, material requirements, safety information, etc. The CMMS triggers PM Works Orders depending on user defined frequency criteria. Multiple triggering is possible. Subsequent PM based Work Orders are created automatically and placed in the Work Order Backlog for execution purpose. The CMMS is equipped with statistical function by maintaining a number of inputs for each entity in a database for analytical use.

A Maintenance, Repair and Operation (MRO) Inventory module is specifically assigned to enable the control of a large number of unique and low-unit value items. The MRO Inventory module automates the reorder process in the CMMS by recognizing calculated safety stock levels, replenishment lead times and sophisticated "available-to-promise" logic based on expected receipts and issues. The MRO Inventory module provides the ability to uniquely identify and track repairable items and critical parts through serialization so that operation staff have ready access to the inventory information and management decision such as inventory status, location, quantity in stock, supply lead time, etc.

The CMMS also generates reports to facilitate operation staff on plant performance monitoring and statistics review by automatic reports generation mode to be scheduled by time of day, day of week, hour of day, or at the end of a shift, or on demand by operation staff. Typical CMMS report contains information on plant influent, plant effluent and dewatering system, their duty/standby train/units status, actual retention time of the units, chemical inventory as well as their summaries on alarms and maintenance activities including corrective and preventive maintenance work orders cleared and not yet cleared. Operation staff can add snapshots of trends, histograms using graphic templates in their CMMS reports.

4 Condition-Based Maintenance Approach

Condition-based maintenance (CBM) has been widely adopted in the industry due to its maintenance efficiency and flexibility [7]. CBM is based on using real-time data to prioritize and optimize maintenance resources. However, modelling, maintaining and using the required data from asset management to implementation of CBM is complex and intensive in nature. Data acquired must be accurate because

it forms the basis for the decision support tools. In fact, CBM is usually performed based on an assessment or prediction of the equipment health instead of its service time so as to reduce down time and enhance operational safety.

Modern diagnostic practice in industry is the combination of human expert knowledge and experience of equipment status and failure with continuously monitoring and analyzing its condition. An effective prognostic model requires performance assessment, development of real-time condition monitoring information and degradation signals, failure analysis, health management and prediction, feature extraction and historical knowledge of faults [8] which necessitate tremendous effort, resources and expertise with every data acquisition. For Upgraded PPSTW, extensive instrumentation of equipment together with computer tools for analyzing and predicting equipment's remaining useful life (RUL) are used in the process of CBM implementation. Online condition monitoring is adopted for applicable equipment to estimate the RUL and achieve maintenance optimization. Accurate prediction of equipment's RUL is somehow critical for its operation and productivity.

For demonstration, taking sewage pump as an example, the first condition is the running hour of the pump with triggering condition referred to the equipment O&M manual. The second condition is operating current of the pump. These conditions are very useful for assessing general loading status of the pump. The CMMS monitors all the duty and standby pumps to record their running conditions. Operation sequence is assigned with priority to the loading conditions and health status of each pump instead of traditional averaging running hour approach. The last condition is the establishment of Degrading Index for equipment with RAR model for the forecast of RUL of the equipment. Above mentioned three conditions are implemented simultaneously in the CMMS in determining maintenance schedule and works planning. In order to facilitate the operation of SCADA and CMMS, CMMS client desktop is integrated into the SCADA system as shown in Fig. 3. In the CMMS, operation staffs are able to monitor, operate and schedule

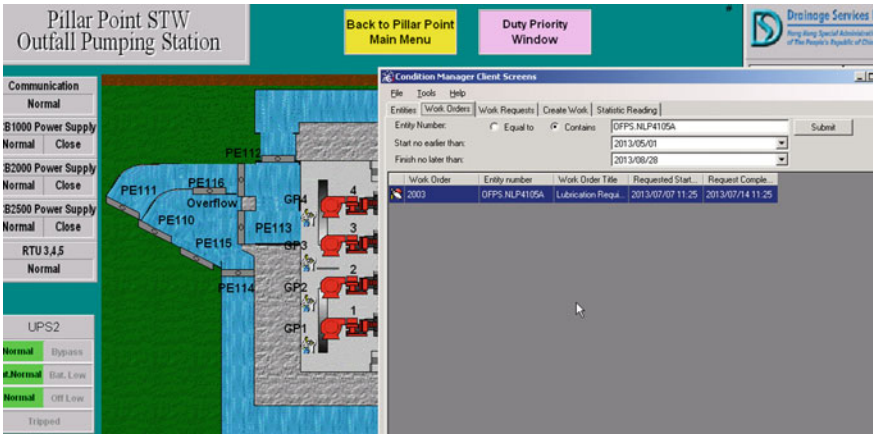


Fig. 3 SCADA interface with CMMS desktop integrated

maintenance with single screen user interface and review all particulars of the equipment, Work Request details, Work Order details and generation, resources allocation and scheduling, etc.

5 Case Study—RUL Estimation of the Existing Outfall Screw Pump Bearing

The accuracy of the RUL estimation plays a critical role for the CMMS. As part of testing and commissioning program for the CMMS, a case study has been carried out to examine the work flow for optimal time in generating work request using Degrading Index and Recursive Auto-Regression (RAR) model.

In this case study, the RUL of the existing Outfall Screw Pump bearing is used to compare the Reasonable Time for component delivery or overhaul subcontract tendering in the Upgraded PPSTW. Work flow of the RUL modelling is shown in Fig. 4 while schematic and photo of the experimental setup is shown in Fig. 5.

5.1 Collect Condition Monitoring Data

This is the first stage of the Case Study. Three sets of vibration sensors are mounted to the pump shaft of Outfall Screw Pump No. 4 so as to measure the pattern and magnitude of linear and angular vibration when in operation. The measured field

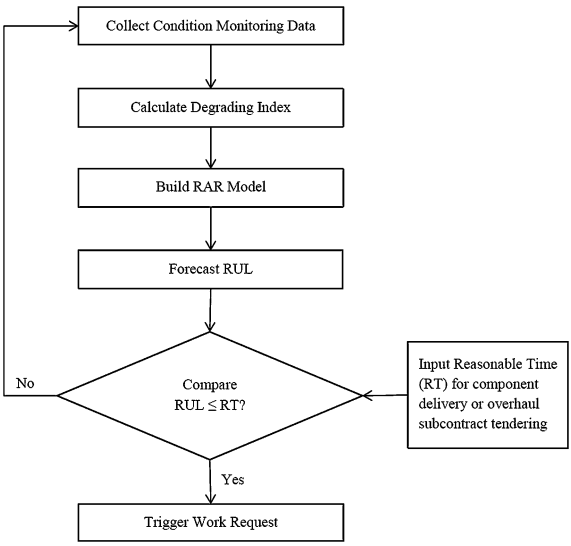


Fig. 4 Work flow of the RUL modelling

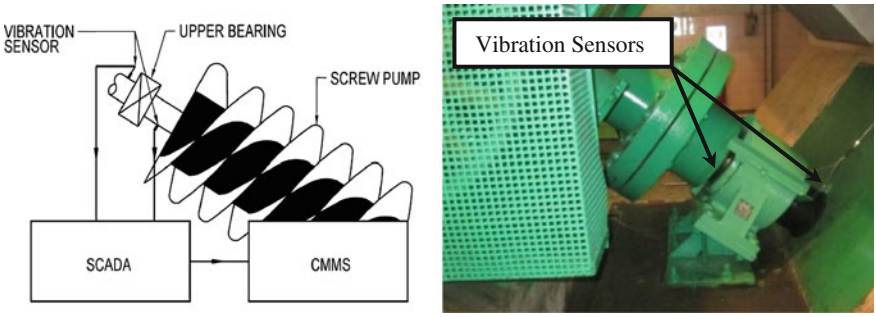


Fig. 5 Schematic and photo of the experimental setup

data is fed to the SCADA system. The condition monitoring data is then retrieved by the condition monitoring engine of the CMMS for subsequent manipulation and processing.

5.2 Calculate the Degrading Index

This is the second stage of the Case Study. For each set of condition monitoring data, Degrading Index is calculated with root-mean-square (RMS) of vibration magnitude.

5.3 Build RAR Model

This is the third stage of the Case Study. The Recursive auto-regression (RAR) model comprises of a time series stochastic autoregressive (AR) model and recursive parameter estimation with an established algorithm based on model order determination with Akaike's Information Criterion (AIC) [9] and the parameters estimate with Recursive Least Square method [10]. The RAR model is then integrated to the condition monitoring engine of the CMMS as one of the rules in the decision support tool.

5.4 RUL Estimation—Degrading Index Forecasting

This is the fourth and the last stage of the Case Study. When the future degrading index forecasted in the time series equation is equal to or more than the threshold degrading index, the time step ahead corresponding to this forecasted degrading index is considered to be RUL of the machine. The RUL is compared with

Reasonable Time (RT) for component ordering and delivery or subcontracting overhaul works. If RUL is still longer than RT, then the condition monitoring process will continue and proceed as no follow-up action is anticipated. However, when the RUL is equal to or less than RT, a Work Request for Standard Overhaul Procedure will have to be initiated.

6 RUL Modelling System Test

Before the implementation of RUL prediction in practical application, a numerical simulation of the model is performed to assess its accuracy and triggering function. In this practical application, the senor and transducer will take vibration measurement and send the measured signals to SCADA system and CMMS one by one in time sequence. In the system test, a signal generator is employed to play the role of sensor and transducer. It generates a series of signals to SCADA system and CMMS according to a preset governing equation plus ±1 % random part. The series of signals simulates vibration data and the degrading indexes.

During the test, the CMMS system automatically built a RAR(5) degrading index model according to the received data from the signal generator while 5th order model had a minimum AIC value shown in Fig. 6. RAR(5) model was adopted to predict when the future degrading index was equal or nearest to preset threshold degrading index of 3,000. For the result shown in Fig. 7, the degrading index of 3061 was reached at 90th step which was forecasted at the 86th step. Time for four steps ahead forecasting (90th–86th) was 8 weeks as 2 weeks was for each

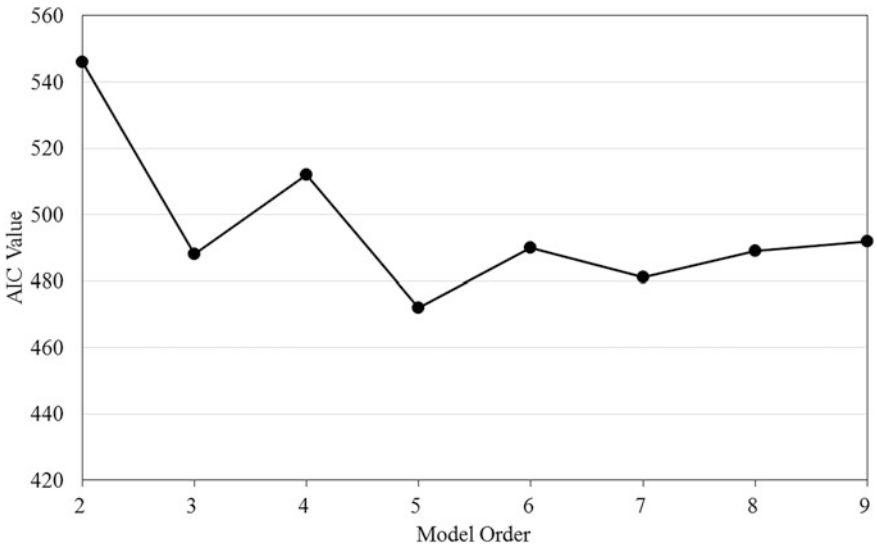


Fig. 6 AIC value with different order of model

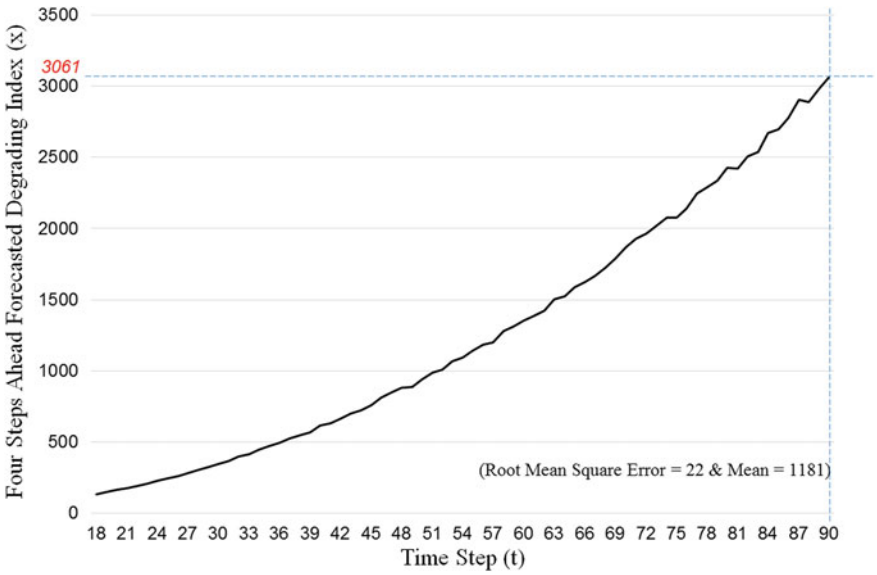


Fig. 7 Four steps ahead forecasted degrading index

sampling time step. RUL was considered as 8 weeks which was equal to the preset RT of 8 weeks. As such, work requested was triggered automatically at 86th step to allow 8 weeks for component delivery and overhaul subcontracting. Also, the root mean square error of 22 and the mean of 1181 were found based on the following Eq. (1). The error-to-mean ratio of 1.86 % (i.e. 22 ÷ 1181) comprised of two parts with one part an accumulated inaccuracy of multiple forecasting steps and the other part of ± 1 % pure random and uncorrelated in each raw signal is considered acceptable.

$$Root\ Mean\ Square\ Error = \sqrt{\frac{\sum_{i=t}^{N}(X_i - \overline{X})^2}{N}} \tag{1}$$

where X_i is the ith set of actual degrading index originated from the signal generator, X is the forecasted degrading index from RAR(5) model built in CMMS and N is number of signal.

Using the Eq. (2), the forecasting accuracy for degrading index of 99.18 % was found at the time step of 90. In this case, the degrading index is vibration magnitude that is likely possible to indirectly represent the wearing of pump shaft's bottom bush which is submersed in sewage during operation. The inaccuracy of 0.82 % for predicting the bush's wearing is acceptable in industry.

$$\text{Forecasting Accuracy}\,\% = \left(1 - \frac{\sqrt{(X_i - \overline{X})^2}}{X_i}\right) \times 100\% = \left(1 - \frac{\sqrt{(3036 - 3061)^2}}{3036}\right) \times 100\,\% = 99.18\,\% \quad (2)$$

7 Conclusion

Though SCADA and CMMS has been adopted in various sewage treatment works and pumping stations of DSD, in this study, condition monitoring with the methodology of RAR for assessing RUL of the equipment is performed and integrated to the Plant's CMMS and SCADA as a part of equipment diagnostic program for the optimization of asset management system. Plant operation, maintenance and resource management is utilized seamlessly under this platform integration. Meanwhile, the proposed RUL forecasting modelling is verified and proved with high accuracy by numerical simulation. This equipment life forecasting model will be further validated by field experimental test and to be implemented for various equipment.

References

1. Fung HK, Cheung SM, Leung TP (1998) The implementation of an error forecasting and compensation system for roundness improvement in taper turning. Comp Indus 35:109–120
2. Institute of Asset Management (2008) Asset management part 1: specification for the optimized management of physical assets. in PAS 55-1, BSI, Editor
3. Baird G (2011) Defining public asset management for municipal water utilities. J Am Water Works Assoc 103:5–30
4. Kumar D, Setunge S, Patnaikuni I (2010) How to develop a practical asset management tool? Proceedings of WCEAM, Springer, London, pp 519–529
5. Maguire D (2008) The business benefits of GIS: an ROI approach. ESRI Press, Redlands, CA, ISBN:978-1-58948-200-5
6. Walt B (2009) Back to basics: SCADA, automation TV: control global—control design. Video clip broadcasted under "Process Automation TV" produced by "Control" and "Control Design". Available at YouTube (http://youtu.be/bfxr5DikdP0)
7. Cong L, Miao Q, Liu Z, et al (2010) Fault diagnosis of gearbox based on interpolated dft with the maximum sidelobe decay windows. Proceedings of WCEAM, Springer, London, pp 133–141
8. Lee J, Ni J, Djurdjanovic D, Qiu H, Liao H (2006) Intelligent prognostic tools and e-maintenance. Comput Ind 57:476–489
9. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Auto Cont 19 (6):716–723
10. Soderstrom T, Ljung L, Gustavsson I (1978) A theoretical analysis of recursive identification method. Automatica 14:231–244

Strategic Asset Management Approach for Sewage Treatment Facilities in Drainage Services Department, the Government of Hong Kong Special Administrative Region

Michael K.F. Yeung, Gary W.Y. Chu and K.Y. NG

Abstract The Drainage Services Department (DSD) is tasked, among other commitments, to operate and maintain (O&M) about 280 sewage treatment facilities, ranging from pumping stations to sewage treatment works at an annual O&M departmental expense amounting to over HK$760 million. To ensure the life cycle cost of the Sewage Electrical and Mechanical (E&M) Assets at these facilities can be optimized, an asset management system based on PAS55-1:2008 standard has been developed and being tried at seven regional plants in the territory. It is aimed that the pilot studies can provide useful insight and solid foundation in establishing an asset management system eligible for PAS55 certification for the pilot plants to be completed tentatively by the end of 2013. This paper introduces the roadmap of developing the Sewage E&M Asset Management System based on PAS55-1:2008 standard and the progress achieved thus far. In the course of its development, the Hong Kong Quality Assurance Agency has been engaged as project consultant to provide asset management studies including feasibility study, staff training, overview of major gaps, framework establishment by enhancement of existing DSD's Integrated Management System, pilot plant maturity review and assessment, etc. The pilot studies mark the evolution of strategic asset management approach in three stages, namely—Stage 1: To establish the scope and objectives of the studies leading to the development of an Asset Management Improvement Plan for Stage 2: To establish a unique asset management system for each of the pilot plants. The effectiveness and efficiency of the asset management system that established can then be determined through pilot implementation for a period of time. Stage 3: To adopt and

M.K.F. Yeung (✉) · G.W.Y. Chu
Drainage Services Department, The Government of the Hong Kong Special Administrative Region, Hong Kong, China
e-mail: mkfyeung@dsd.gov.hk

G.W.Y. Chu
e-mail: garychu@dsd.gov.hk

K.Y. NG
Hong Kong Quality Assurance Agency, Hong Kong, China
e-mail: ky.ng@hkqaa.org

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_78

extend the demonstrated successful examples to all other Sewage E&M Assets of DSD. It is envisaged that the pilot studies will lead to their initial certification of PAS55 for ultimate goal of attaining full certification of all DSD's Sewage E&M Assets within 5 years.

Keywords PAS 55 · Gap analysis · Life cycle cost · Asset register · Gap analysis · Radar chart · Risk assessment · Asset management plan · Level of services

1 Background

The vision of Drainage Services Department (DSD), which was established on 1 September 1989, was to provide world-class wastewater and stormwater drainage services enabling the sustainable development of Hong Kong. In a departmental retreat held in 2011, DSD has developed six strategic goals, one of which (i.e. DSD goal 4(b)) was to establish a Total Asset Management (TAM) system to optimize the long term operation and maintenance of DSD's electrical and mechanical assets.

Being a government department, DSD has been managing substantial amount of drainage infrastructural assets. For the collection of sewage generated from within the territory of a size of about 1,100 km^2, DSD has developed a sophisticated sewerage network with a total length of approx. 1,600 km capable of handling approx. 2.90 million m^3 of sewage every day. There are about 290 sewage treatment facilities ranging from sewage pumping stations to sewage treatment works. The sewerage network is now serving about 93 % of the population in Hong Kong [1].

The Electrical and Mechanical (E&M) Branch of DSD is tasked with the responsibility, among other things, to plan, design, construct, operate and maintain sewage treatment facilities. To maintain this large amount of E&M assets, substantial operation and maintenance (O&M) expenses have been incurred, with an annual departmental expense amount of about HK$ 760 M in E&M works. In this respect, an asset management (AM) system is a useful means to achieve not only an optimal life cycle cost in the long run, but also reliable performance to meet the level of services expected from the general public.

In order to enhance the effectiveness of AM, DSD has set out a roadmap leading to the development of a TAM System in E&M Branch of DSD since 2011. The approach and the progress achieved thus far are elaborated in the following sections.

2 Overview of TAM System

British Standards Institution's Publicly Available Specification 55 (PAS 55) is a well established international practice in AM and is adopted as the framework for the development of TAM system in E&M Branch of DSD. According to PAS 55,

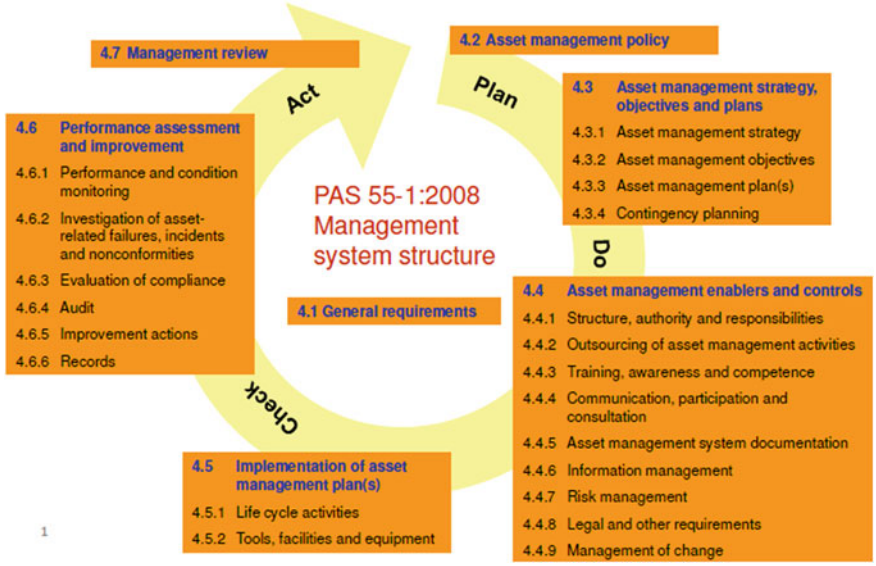


Fig. 1 AM process (*Source* PAS 55-1:2008)

AM is defined as systematic and coordinated activities and practices through which an organization optimally and sustainably manages its assets and asset systems, their associated performance, risks and expenditures over their life cycles for the purpose of achieving its organizational strategic plan. It consists of a Plan-Do-Check-Act cycle with structured risk-based and self-improved management system applied across to achieve optimal life cycle cost. The AM process [2] is illustrated in Fig. 1.

The AM system would be designed primarily to support the delivery of an organizational strategic plan, which in turn aims to meet the expectations from a variety of stakeholders and organizational goals. This provides a coherent direction and guidance from top management to frontline across the whole organization to manage these expectations [3].

The AM system would be guided by AM policy, strategies, objectives and AM plans. These, in turn, direct the different combination of life cycle activities to be applied across a diverse portfolio of asset systems and assets optimally in accordance with their criticalities, condition and performance. Contingency plans are also established for identifying and responding to incidents or emergency situations, as well as maintaining the continuity of critical business activities. The following enabling elements ought to be addressed in the implementation of AM system:

i. structure, authority and responsibilities;
ii. outsourcing of AM activities;
iii. training, awareness and competence;
iv. communication, participation and consultation;

v. AM system documentation;
vi. information management;
vii. risk management;
viii. legal and other requirements; and
ix. management of change.

To facilitate AM implementation, AM plan would be developed to identify the various tasks to be applied for asset life cycle management in order to achieve each AM objective. The plan also defines appropriate responsibility and authority as well as timescales to each individual task, so as to meet the overall timescale of the related AM objective. It also includes long-term asset replacement programmes to provide an overview on asset replacement requirements and associated funding needs in future so that replacement alternatives and expenditure can be planned ahead smoothly.

3 Strategic Goals

A Task Force chaired by a chief engineer was established in December 2011 with members drawn from all 3 divisions of E&M Branch to oversee the development and implementation of AM with the following 3 main initiatives (Fig. 2):

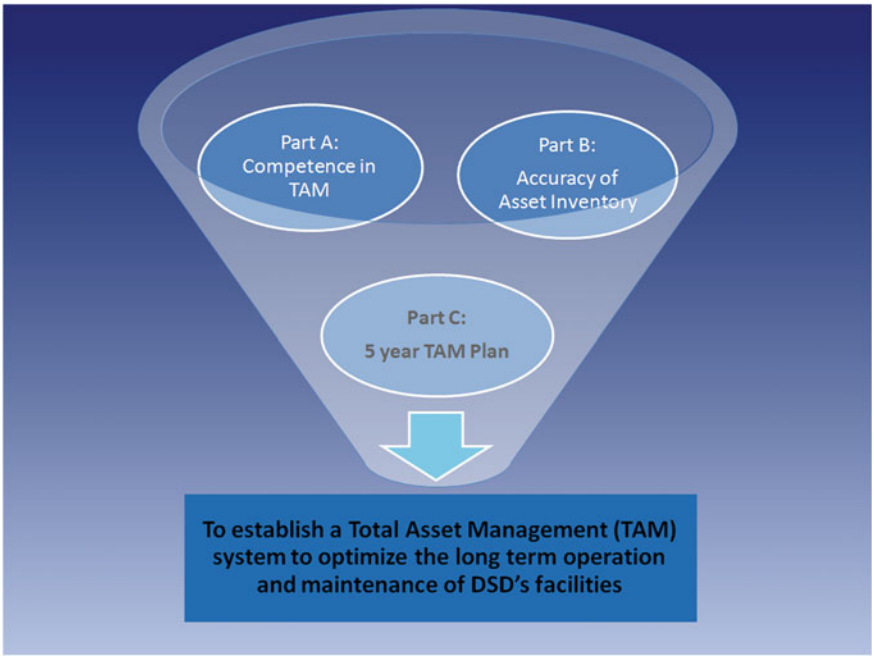


Fig. 2 Strategic AM goal in E&M branch of DSD

Part A: To build up competence in TAM
Part B: To improve the accuracy of asset inventory
Part C: To develop first 5-year TAM plan

Various short, medium and long term action plans were then developed to build the TAM system in stages with their progress monitored by the Task Force. At the next level higher up, the Task Force would report the progress and achievement to DSD's top management on quarterly basis amongst other goals, including goal 4(a) related to DSD's civil assets. These actions plans are briefly elaborated in the following sections.

4 Competence in TAM

In the AM development journey, it is essential to ensure our staff engaged in various activities to possess an appropriate level of competence in terms of education, training and experience. The Task Force therefore strongly supports that training plans should be in place to develop the competence of the staff involved as an on-going initiative. The plan includes both local and overseas training activities. 6 sessions of awareness training were first conducted in July and August 2012 aiming to introduce the rationale in establishing the AM system, the concept of risk based maintenance strategies, AM framework under PAS 55, and respective individual roles and responsibilities, etc. with more than 100 staff trained.

The Task Force then organized visits to various PAS 55 certified utilities locally to exchange views and share their experience since 2011. These included Hongkong China Gas, MTR Corporation, CLP Power, Hongkong Electric, etc. These visits provided invaluable opportunities for our professional and technical staff to acquire a better understanding from these leading companies renowned for their dedication and achievements in AM practices. It was worthy to note that all these companies have extensive experience in implementation of ISO9001 quality management and ISO14001 environmental management systems. These two management systems have many core attributes in common with those of PAS 55 asset management system. It is considered that this is a key successful factor in launching PAS 55 asset management system. The other factor is effective communications in different forms during the AM journey.

Selected DSD staff also attended overseas AM training courses/conferences in Australia (2010), United States (2011), South Korea (2012) and United Kingdom (2013) to keep abreast on the best AM practices as well as practical skills and knowledge. In addition, these visits provide valuable opportunities for DSD to cultivate the networking with other AM professionals.

As part of DSD's effort to promote continuous professional development, intermediate level AM training courses such as PAS 55 internal auditor training, risk assessment workshops, etc. would be launched in late 2013 to equip relevant staff members with essential skills to conduct internal audit.

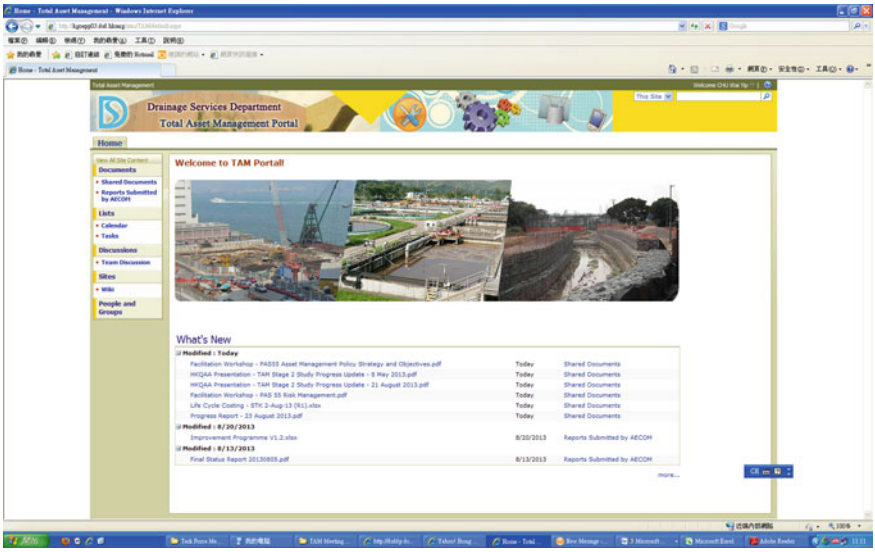


Fig. 3 TAM portal in DSD for information sharing

One of the key successful factors in implementation of AM system is effective communication. A TAM portal (Fig. 3) was therefore established in April 2012 to share pertinent AM information and knowledge within DSD. All the training materials, duty visit reports, conference synopsis, consultancy reports and TAM Task Force meeting minutes, etc. have been uploaded to the portal for knowledge sharing.

5 Asset Inventory

Good AM system requires meaningful, quality, timely AM information for support. Currently there are two types of computerized maintenance management system (CMMS) in E&M Branch's sewage treatment divisions, namely, MAXIMO®[1] and Advantis Pro®[2] to support this initiative. As 2 different types of CMMS are being used in various sewage treatment facilities, this arrangement has introduced system scalability problems and has caused problems in data integration. DSD commenced a feasibility study for implementing a solution that could resolve data integration and system scalability problems. The study recommended implementing an integrated Sewage Treatment Operation and Maintenance Management Information System (STOMMIS).

[1] Pilot Sites for PAS55 Implementation – Ho Pong Street Sewage.

[2] Advantis Pro system applied in regional centres such as Shatin, Sai Kung, Taipo, Shek Wu Hui, Yuen Long, Sham Tseng, etc.

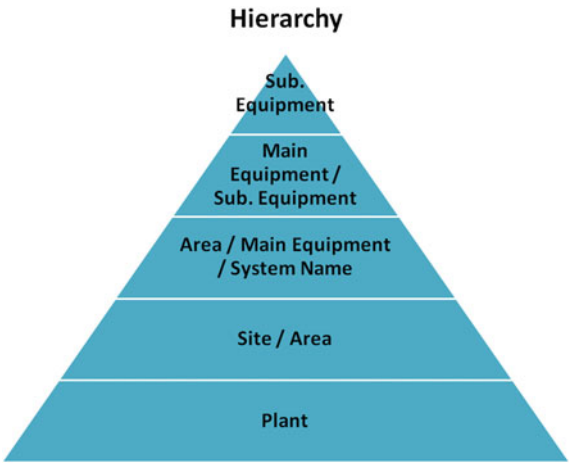


Fig. 4 AR and its hierarchy format

The proposed system includes the deployment of a central information repository to facilitate the storage, updating and retrieval of operational and maintenance information as well as the centralized management report for useful asset life cycle information such as asset capacity, cause of failure, maintenance downtime, etc. Data collected by STOMMIS would be stored in a data warehouse for analysis to support optimization of maintenance activities and the development of asset replacement plan based on asset conditions and risks. The current CMMS system has been progressively migrating to integrate with STOMMIS which is targeted to complete by end 2013.

A good CMMS should embrace a clear identification and definition of asset items that will be managed during the asset life cycle. As such, AR and its hierarchy format have been standardized so that assets data can be effectively stored, retrieved and manipulated by the users. All assets are named according to a pre-defined hierarchy which include name of plants, sites/areas, main system names, main and sub-equipment names (Fig. 4). We have been making good progress in implementing the standardized AR in major regional centres.

The CMMS was designed to capture the AM information and make them readily available to all designated users. In the daily O&M of sewage treatment facilities, however, such information including equipment maintenance history data, causes of failure, stock of spare parts, etc. are stored in CMMS but may not be readily available to frontline staff out in the field to make maintenance decision promptly. To enhance the maintenance decision efficiency, a trial project on application of quick response (QR) code system was successfully launched at Sham Tseng Sewage Treatment Works in October 2012. The system enables frontline staff to retrieve useful asset information on site using mobile devices such as smart phones and tablets (Fig. 5). The effectiveness of the QR code system would be reviewed in April 2014.

Substantial effort has also been made to maintain the asset inventory information through periodic audit and review. To improve the accuracy of inventory and streamline the process, a trial project on application of radio-frequency identification

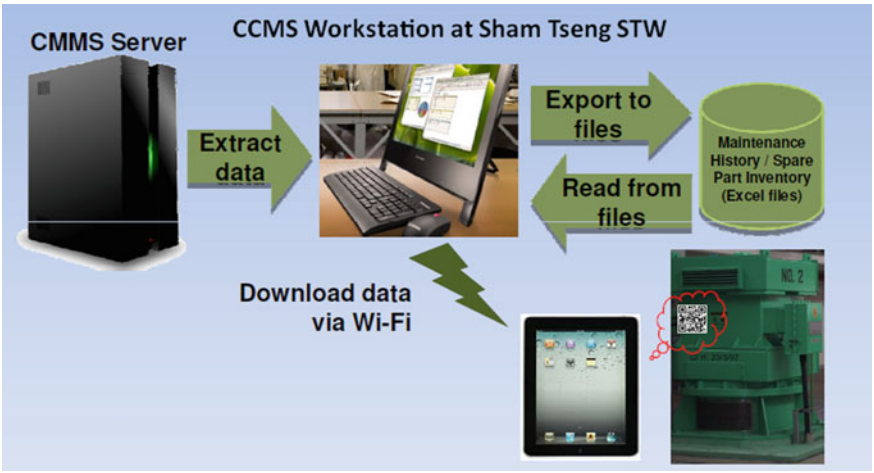


Fig. 5 Application of QR codes in up-keeping maintenance history of E&M Assets

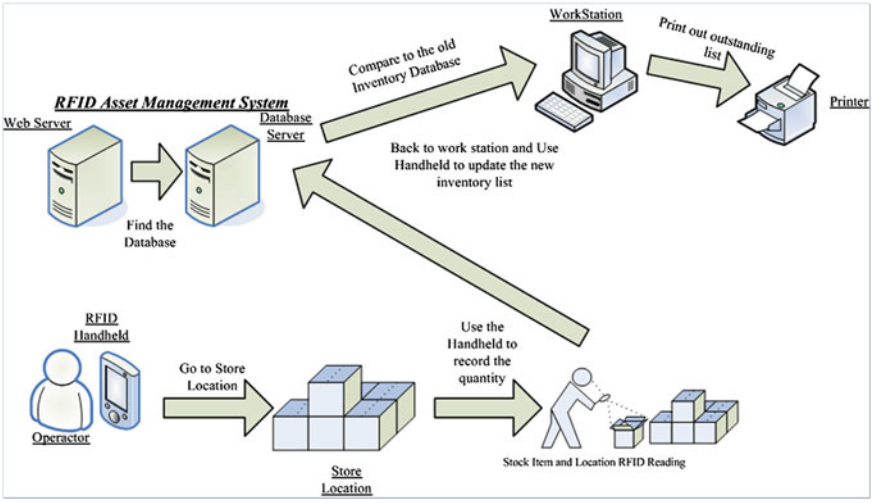


Fig. 6 Application of RFID codes in inventory control

(RFID) system was successfully launched at Shatin Sewage Treatment Works in October 2012. All the spare parts in store had been attached with a RFID tag which can be readily detected wirelessly making use of radio-frequency electromagnetic fields by a portable device (Fig. 6). This significantly enhances the efficiency in stock-taking and the accuracy of stock record. The effectiveness of the RFID system would be reviewed in April 2014.

6 The First 5-Year TAM Plan

Implementation of TAM system in E&M Branch has been proceeding in stages. A pilot study was first launched in Sewage Treatment Divisions since 2010. 2 number of sewage pumping stations, namely Ho Pong Street sewage pumping station (HPSSPS) at Tuen Mun district and Hung Hom Bay sewage pumping station (HHBSPS) at Hung Hom district (Figs. 7 and 8) were selected as pilot sites. Hong Kong Quality Assurance Agency (HKQAA) was engaged to carry out a gap assessment. The purpose of the study was to determine the gaps between the current AM systems applied in these 2 numbers of sewage pumping stations against the requirements of PAS 55. The following major gaps were identified [4]:

Fig. 7 Pilot sites for PAS55 implementation—Ho Pong street sewage pumping station

Fig. 8 Pilot sites for PAS55 implementation—Hung Hom Bay sewage pumping station

i. AM policy, strategies, objectives and plans consistent with the departmental strategic plan should be established.
ii. A risk management system should be established and applied to enable the proper risk identification and assessment of these assets and asset systems.
iii. The current contingency plans, operational procedures for control of activities across the whole asset life cycle should be reviewed against the results from risk assessments.

Following this exercise, the sewage treatment divisions took it further in enhancing the existing AM system with reference to the recommendations by HKQAA. Action plans were then developed and implemented in 2011 to close the gaps. An audit was conducted in early 2012 to verify the adequacy and effectiveness in closing the gaps previously identified. It was then verified that most of the gaps were generally addressed. Control of asset life cycle activities including the utilization and maintenance of assets were found generally in good order.

The above pilot study provided useful insight and solid foundation in establishing an AM system targeted for PAS 55 certification at a later stage. It marked the evolution of strategic TAM system development in 3 stages, namely:

Stage 1: To establish the scope and objectives of a comprehensive study leading to the development of an AM Improvement Plan (AMIP) at E&M Branch level.

Stage 2: To establish an AM system at each Selected Critical Plant (SCP). The adequacy and effectiveness of the AM system thus established can be determined through pilot implementation for a period of time.

Stage 3: To populate the established AM system from Stage 2 progressively to other sewage treatment facilities in E&M Branch of DSD.

Various stages of TAM system development is illustrated in Fig. 9.

HKQAA was engaged as project associate in Stage 1 and Stage 2 to provide necessary training and gap assessment to aid DSD in the development of TAM System including feasibility study, staff training, overview of major gaps, framework establishment by enhancement of existing DSD's Integrated Management System (IMS), pilot plant maturity review and assessment, etc. The major works at each of these 3 numbers of stages are elaborated as follows:

6.1 Stage 1 (October 2012–February 2013)

HKQAA conducted a desk-top study across the full asset portfolio in E&M Branch to identify critical plants for detailed gaps assessment. To facilitate the selection of critical plants, a data collection form (Fig. 10) was developed for individual Works Managers to complete so that relevant AM information for each DSD owned sewage treatment works can be collected.

Following that, all the plants were rated into 4 levels in the aspects of cost, performance and risk with the aids of statistical methods. Level 1 is the bottom 25 % of the total DSD owned sewage treatment works with least significance. Level

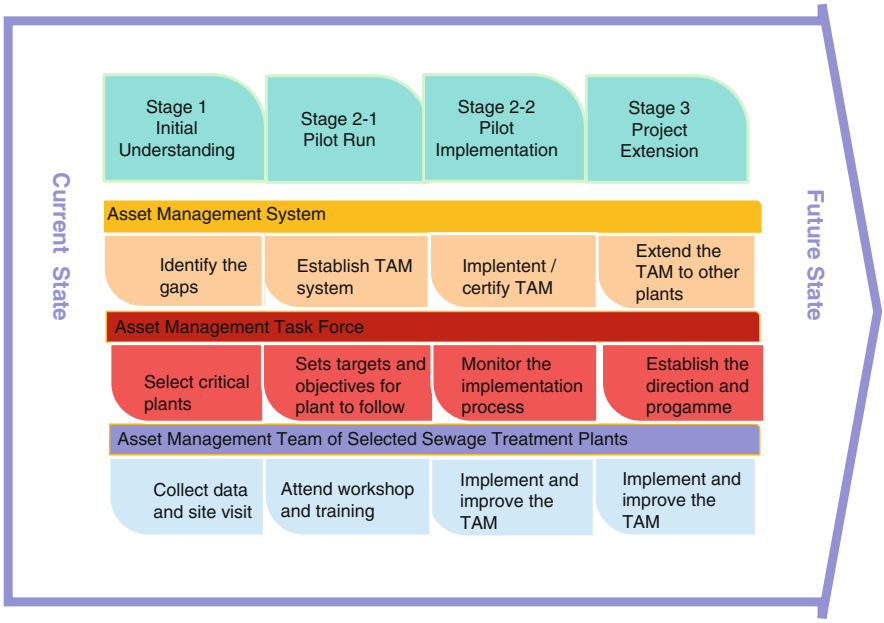


Fig. 9 Development stages of TAM system

Data Collection Form – Stage 1 Asset Management Studies for DSD Owned Plants under DSD Goal 4b

Division:	☐ ST1 ☐ ST2			
Regional Control Centre:	Name:	e.g. Sham Tseng and Tuen Mun Region		
	DE Cost in FY2011/12 (HK$):		Establishment (no. of staff) strength:	
Name of Sewage Treatment Works:		Handover Year:		
Design Flow (m^3/day):		Average Flow in FY11/12 (m^3/day):		
Treatment Process:	☐ Chemically Enhanced Primary Treatment ☐ Preliminary Treatment ☐ Secondary Treatment ☐ Tertiary Treatment		☐ Rotating Biological Contactors ☐ Oxidation Ditch ☐ Sequencing Batch Reactor ☐ Others (please specify):	
No. of Associated Upstream Sewage Pumping Stations:		No. of non-compliance to EPD discharge license in FY 2011/12:		
Type of Computerized Maintenance Management System	☐ AdvantisPro ☐ Others (please specify): ☐ Maximo			
Asset Replacement and Upgrading Cost (HK$):	Capital Account Items commenced in FY11/12:			
	Cat. D Items commenced in FY11/12:			
Total Term Maintenance Contract Cost (civil + E&M) in FY11/12 (HK$):				
Total No. of CM Works Orders in FY 11/12:				
No. of Technical Complaints:				
System Availability (%):				
Emergency Electricity Backup:	☐ Dual feed from power company ☐ Emergency generator (fixed installation) ☐ Uninterruptable Power Supply for SCADA ☐ Others (please specify):			

By Works Manager

Name:
Date:

Form TAM001

Fig. 10 Data collection form

Plants	1. Normalized Staff	2. Normalized Maintenance Cost (M)	3. Normalized Asset Replacement and Upgrading Cost (M)	4. No. of CM Works Orders (normalized)	5. No. of Technical Complaint	6. System Availability (%)	7. Extent of Electricity Back up	Total
Sai Kung STW	4	4	1	4	1	1	3	18
Sham Tseng STW	4	4	1	3	1	1	3	17
Sha Tau Kok STW	4	3	1	4	1	1	3	17
Shek O PTW	4	3	1	3	1	1	3	16
Siu Ho Wan STW	3	4	1	3	2	1	1	15
Stonecutters Island STW	1	2	1	1	2	1	1	9
TO Kwa Wan PTW	1	2	1	1	1	1	2	9
San Wai Preliminary Treatment Works	1	1	1	1	1	1	3	9
Central PTW	2	2	1	1	1	1	1	9
Wan Chai East PTW	2	2	1	1	1	1	1	9

Fig. 11 Selection of critical plants for detailed assessment

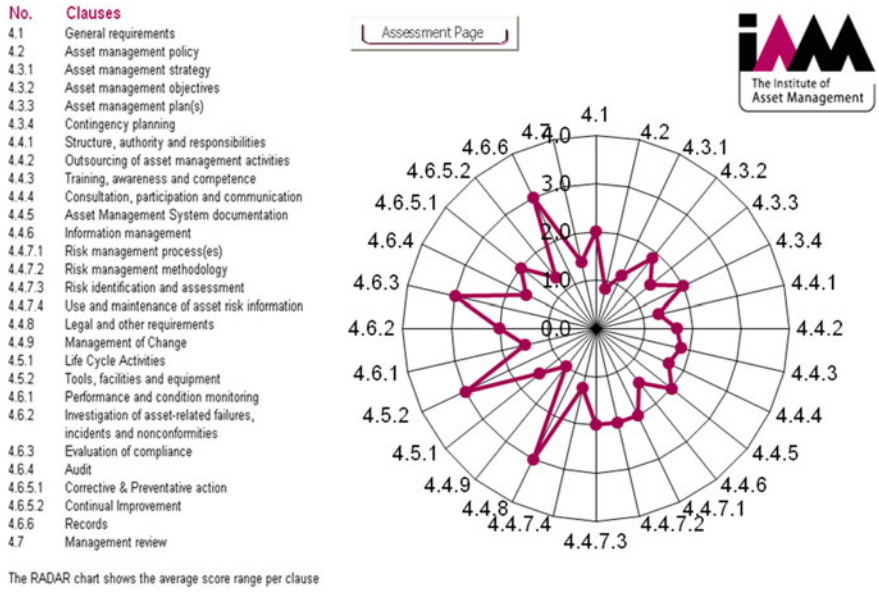


Fig. 12 RADAR chart

2 and 3 are the two respective middle 25 % ranges with moderate significance whereas level 4 is the top 25 % with the most significance. The total ratings of all plants were then compiled and arranged in descending order. The top 7 numbers of plants with higher total ratings were then shortlisted accordingly as SCP for further assessment at Stage 2 study (Fig. 11).

The current AM practices of the 7 SCPs were then holistically reviewed using the assessment tool developed by the Institute of Asset Management to determine the maturity level of current AM practices [5]. The results were presented in the form of RADAR chart to benchmark the average score range per PAS 55 clause (Fig. 12).

This assessment provided a practical direction for improvement to enhance the current AM system in order to close the gap to meet PAS 55 requirements. An Asset Management Improvement Plan (AMIP) [6] was then established. It summarized the findings and lesson learnt from on-site survey, identified the overall strengths and weaknesses, the major gaps which existed and presented a prioritized plan to raise the maturity level of the SCPs in the next 15 months.

6.2 Stage 2 (March 2013–May 2014)

The Stage 2 study was aimed to materialize all the recommendations in AMIP by launching pilot projects in the SCPs so as to establish AM system eligible for PAS 55 certification. The 7 numbers of SCPs (Fig. 13) together with the pilot sites at HPSSPS and HHBSPS were identified for implementation of TAM system in Stage 2. The Study commenced in March 2013 for completion by May 2014.

During Stage 2 of the AM system development, HKQAA organized a series of AM training courses and on-site facilitation workshops pinpointing the gaps of the SCPs to equip Nominated Plant Representatives (NPRs) with the skill sets to

Fig. 13 Location plan for SCPs

establish all the required TAM systems, documentation, procedures, etc. as necessary for trial implementation. NPRs were professional engineers and technical staff from various disciplines nominated from each SCP. It is the prerequisite that NPRs should be familiar with the existing operation, processes, controls, instructions and asset performance, etc. The NPRs were then facilitated by HKQAA to go through the various critical elements of an AM system.

The facilitation workshops provided a platform for NPRs to have an in-depth understanding and overview of the current AM operation, ranging from organizational strategies, through planning, operational processes and control, to existing and desired performance. This is an important process for gathering and consolidating all the useful and valuable knowledge from our experienced technical colleagues. And throughout the process, the NPRs focused on a number of elements that would be crucial in establishing an AM system, these included but not limited to establishing the following:

i. AM Risk Framework
ii. AM Database
iii. AM Plan
iv. TAM Development Plan (TAMDP)

6.2.1 AM Risk Framework

The NPRs first defined the boundary of the AM system by consolidating an asset registry to contain all the physical assets at each SCP. NPRs were then guided to holistically review the criticality of individual equipment. Depending on the functional importance, equipment was categorized into the following critical or non-critical assets (Fig. 14).

Type of Equipment	ID	Name of Asset	Location	Critical / Non-critical	Asset Condition
Inlet Works					
Inlet Penstock	HP -SPS-01-016	Motorised Penstock No.1 (Control Panel)	Inlet Works	N	2
	HP -SPS-02-004	Motorised Penstock No.1	Inlet Works	C	2
	HP -SPS-02-005	Motorised Penstock No.1 (Actuator)	Inlet Works	N	2
Bypass Penstock	HP -SPS-01-017	Motorised Penstock No.2 (Control Panel)	Inlet Works	N	2
	HP -SPS-02-008	Motorised Penstock No.2	Inlet Works	C	2
	HP -SPS-02-009	Motorised Penstock No.2 (Actuator)	Inlet Works	N	2
Screenings	HP -SPS-02-018	Basket Screen	Inlet Works	C	1
	HP -SPS-02-051	Gripper	Inlet Works	C	3
Outlet					
Outlet Penstock	HP -SPS-02-012	Manual Penstock No.1	Pump House	N	2
	HP -SPS-02-013	Manual Penstock No.2	Pump House	N	2
	HP -SPS-02-014	Manual Penstock No.3	Pump House	N	2
	HP -SPS-02 6727	Manual Penstock No.4 (Before Flow meter)	Pump House	N	2
	HP -SPS-02 6728	Manual Penstock No.5 (After Flow meter)	Pump House	N	2
Outlet Check valve	HP -SPS-02-015	Check Valve No.1	Pump House	C	1
	HP -SPS-02-016	Check Valve No.2	Pump House	C	1
	HP -SPS-02-017	Check Valve No.3	Pump House	C	1
Pump	HP -SPS-02-001	Sewage Submersible Pump No.1	Pump House	C	1
	HP -SPS-02-002	Sewage Submersible Pump No.2	Pump House	C	2
	HP -SPS-02-003	Sewage Submersible Pump No.3	Pump House	C	2

Fig. 14 Identification of critical assets

Risk Assessment Matrix					
Likelihood (see Note 1)	Consequences (see Note 2)				
	Insignificant (1)	Low (2)	Moderate (3)	High (4)	Hazard (5)
Rare (1)	L	L	L	L	M
Unlikely (2)	L	L	M	M	M
Possible	L	L	M	H	H
Likely (4)	L	M	H	H	VH
Often (5)	L	M	H	VH	VH

Risk Rating		Action Required
VH	Very High Risk	Immediate corrective action
H	High Risk	Prioritized action required
M	Moderate Risk	Planned action required
L	Low Risk	Managed by routine procedures

Fig. 15 Risk assessment matrix

i. Critical assets (C)—vital to system operation and has certain impact to system operation in case of breakdown/failure.
ii. Non-critical assets (N)—no impact to normal system operation in case of breakdown/failure.

The NPRs then conducted a risk assessment for all the critical assets with due consideration of the consequence and impact of the risk events in aspects such as system failure, public confidence, legal and financial issues as well as the likelihood of occurrence of the risk events. A risk assessment matrix was thus developed and used to prioritize the risk level of all the critical assets across their life cycle (Fig. 15).

For those critical assets at risk, proper risk treatment plans were established to manage the relevant risk events (Fig. 16). In addition, past performance and failure data were also retrieved and consolidated for failure causes analysis to derive the corresponding mitigation measures of the overall risk management plan.

This aforementioned review allows NPRs to re-visit holistically their physical assets in terms of the current condition, performance and risk levels in a systematic manner so that maintenance activities for critical assets can be prioritized sensibly. In addition, review on past performance and failure data could help to realize

Asset / Asset Group	What can Happen / Risk Event	Possible Effect	Inherent Risk			Existing Control	Risk treatment plan
			L	C	Rating		
The entire SPS	Power suspended by CLP	sewage bypass, upstream catchment	1	5	M	Generator set backup	Refer to HPS Contingency plan
	Term contractor not performs	Delay in operation and maintennce work	1	1	L	Use direct labour to undertake operation	No additional control
	Flooding	Sewage overflow, damage to equipment	1	2	L	Follow existing procedure	Refer to HPS Contingency plan, provide sufficient spare part
	Blockage of access road	Delay disposal of screening	1	1	L	NA	Provide sufficient spare container
	Strong typhoon	Damage to facilities	3	1	L	Follow existing procedure	Refer to HPS Contingency plan
	Fire	Causing injury	1	5	M	Follow existing procedure	Refer to HPS Contingency plan
		Damage to facilities	1	5	M		
	Outbreak of highly contagious disease	Staff infected by disease	1	3	L	follow instruction from the government	Refer to instruction from the government

Fig. 16 Sample of risk treatment plan

objectively the likelihood and impact of the occurrence of various risk events so that effective monitoring and control measures could be planned in the management system.

6.2.2 AM Database

Having identified the critical assets with moderate risk (M) or above, NPRs further undertook to review the performance requirements, current asset condition, failure patterns, cause of failures, etc. Based on the historical corrective maintenance records, NPRs tried to correlate the relationship between potential failure events and their pre-failure symptoms (Fig. 17).

Having consolidated all the pre-failure symptoms of these critical assets, NPRs established a data collection mechanism to capture all the required information in CMMS for on-going monitoring the performance and trend-to-fail of critical assets. With this mechanism in place, potential failure can be minimized by rectification at pre-failure stages. These valuable leading performance indicators such as flow rates, electricity consumption, service availability of critical equipment, etc. can be very helpful tools for prioritizing maintenance effort in managing overall performance and risk of critical assets (Fig. 18).

6.2.3 AM Plan

The AM policy plays a leading role in driving the whole AM system. As such, HKQAA conducted workshops in the Task Force meetings to walkthrough with members the following AM policy, which elaborated the principles, approach and expectations of the AM system.

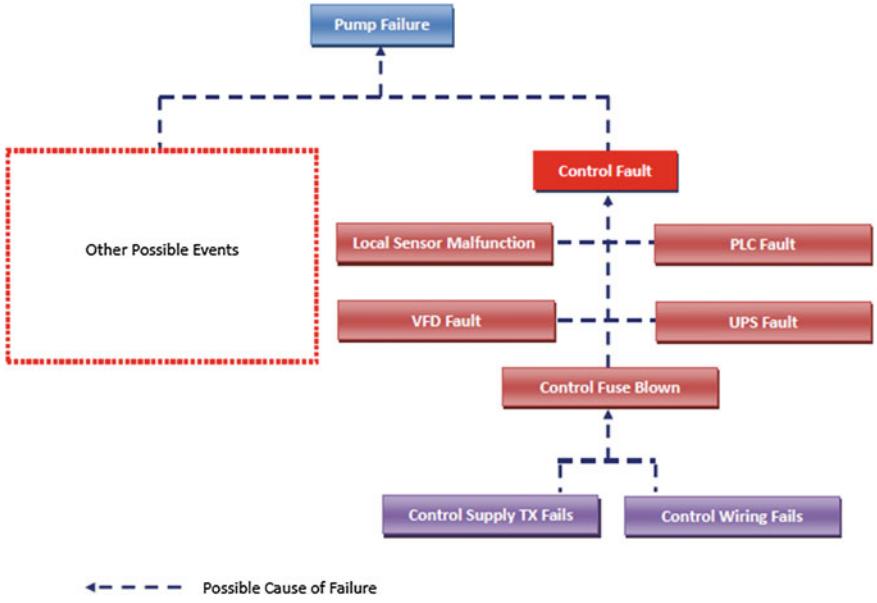


Fig. 17 Sample of failure cause analysis

AM Policy: *DSD Sewage Treatment Divisions are committed to providing a cost effective and reliable sewage treatment service for the public in the following aspects-*

i. *Full compliance with applicable legal requirements.*
ii. *Evaluating the effectiveness of our AM system for continuous improvement and satisfying the needs of our stakeholders.*
iii. *Continuously optimize the cost management, operational performance, risk management and the total life cycle cost, and improve the efficiency and effectiveness of AM processes. To facilitate such, DSD Sewage Treatment Division will maintain current, relevant & reliable data & records of critical assets; as well as understanding & forecasting asset maintenance & renewal costs.*
iv. *Regular review of the AM Strategy, Objectives and Plans*

Along with this AM policy, NPRs have established corresponding AM plan that covered the management strategies and action plans of the critical assets across their life cycle, from acquisition, utilization, maintenance, to disposal. To prepare the AM plan, NPRs consolidated the asset life cycle information including demand forecast, asset current condition /performance, asset remaining service life, risk level, acquisition cost, utilization and maintenance cost, failure history, etc. from the CMMS. The following AM objectives would then be determined based on the desired level of services with due consideration to cost, risk and performance:

Item no.	Asset	Failure symptoms / Causes	Severity	Follow Up Items	Remarks
1	Gripper ID No.: HP -SPS-02- _____	□ Unusual noise		□ Change wear parts :_____	Corrective Maintenance Down Time: _____Days
		□ Un-smooth transmission in gear box		□ Change gear oil	
		□ Rake teeth not operated		□ Clear blockage	
		□ Overload			
		□ Loose parts: _____		□ Others_____	
		□ Others: _____			
		□ Normal			
		□ Overheat		□ Change gear oil	
		□ Excessive vibration		□ Clear blockage	
		□ Air lock			
		□ Level sensor fault		□ Others_____	
		□ Computer fault			
		□ Unusual flow			
		□ Other_____			
		□ **Normal operation**			

Fig. 18 Sample data collection form for monitoring performance of critical assets

i. Reduction of corrective maintenance man hours (Expenditure)
ii. Meeting zero overflow incidents resulting from equipment breakdown. (Level of Performance)
iii. Maintaining service availability of selected critical asset to the certain percentage as specified in AM Plan. (Risk)

The AM plan could also embrace life cycle cost management capability which would enable projection of the funding required to meet the levels of performance. It also provides an overview of future asset replacement requirements in a 5-year plan so that replacement alternatives and expenditure profile can be planned ahead.

It is remarked that DSD has implemented an IMS to safeguard its sewage treatment operations. The developed AM system shall be able to link up and handshake with the existing IMS so that maximum synergy between the two systems can be achieved without duplication of effort from operation and system administration point of view.

6.2.4 TAMDP

When the implementation of the AM system in the SCPs becomes mature, a TAMDP which will consolidate the findings and practical experience learnt from the Stage 2 study such as the major hurdles, overall strengths and weaknesses of the organization, etc. will be prepared. TAMDP will present a prioritized resources plan to raise the overall maturity level of E&M Branch of DSD in compliance with PAS 55 in the next 5 years.

6.3 Stage 3 (June 2014–December 2018)

Upon completion of Stage 2 study, the established TAM system shall have meaningful representation of the major types of sewage treatment facilities in E&M Branch of DSD. The similar TAM System can then be populated progressively to other sewage treatment facilities from 2014 to 2018 according to the recommendation from TAMDP. Surveillance audits will also be conducted annually to ensure that the TAM system is implemented in compliance with the requirements of PAS 55.

7 Conclusion

PAS 55 system can provide a holistic and effective framework, which would guide and enable leading public utilities in many parts of world, including those in Hong Kong, to manage their assets and asset systems in a well-balanced (i.e. cost, risk and performance) and self-managed (Plan-Do-Check-Act) settings in order to optimise the whole life cycle cost. The insight on asset management as conceived above was based on the experience gained from overseas duty visits and experience sharing with fellow asset management practitioners locally.

As a government department, DSD is striving to establish a Total Asset Management system to optimize the long term operation and maintenance of DSD's facilities. In line with this departmental goal and sharing the insight mentioned above, a strategic asset management approach for sewage treatment facilities in DSD has been developed. Under strategic goal 4(b), actions to build up competence in TAM, to improve the accuracy of asset inventory and to develop the first 5-year TAM plan have been progressively deployed since October 2012. These studies are still on-going till completion of its stage 2 by May 2014.

In the AM journey, our focus has been placed to have frequent communications with stakeholders involved from the frontline staff to top management. These are manifested by different forms and levels of meetings, training courses, briefing sessions and facilitation workshops to sustain the momentum, which is a key successful factor in this change management process.

The effectiveness of the trial TAM system will be evaluated by April 2014 for refining the ultimate goal to attain full certification of all DSD's Sewage E&M Assets in the next five years as well as the effectiveness on the use of QR code and RFID on retrieval of equipment maintenance history by mobile device and asset inventory information by wireless detection technology respectively.

Acknowledgements The authors would like to express our gratitude and acknowledgement for the contributions from the following parties and opportunities for experience sharing so far in the course of developing the asset management system for E&M Branch of DSD. (i.) Senior directorate of DSD, (ii.) DSD Goal 4(b) Task Force members, (iii.) NPRs from E&M Branch of DSD, (iv.) Facilitators from HKQAA, (v.) MTR Corporation, (vi.) Hongkong China Gas, (vii.) CLP Power, (viii.) Hongkong Electric, (ix.) Hong Kong Airport Authority.

References

1. Drainage Services Department. Drainage Services Department in Brief 2012-13
2. British Standards Institution. PAS 55-1:2008 Part 1: Specification for the optimized management of physical asset
3. British Standards Institution. PAS 55-1:2008 Part 2: Guidelines for the application of PAS 55
4. Hong Kong Quality Assurance Agency. Drainage Services Department Sewage Pumping Station at Ho Pong Street, Tuen Mun and Hung Hom Bay PAS 55-1 2008 Gap Assessment Report
5. Institute of Asset Management *PAS55 Assessment Methodology*
6. Hong Kong Quality Assurance Agency. Drainage Services Department Electrical and Mechanical Branch PAS 55-1 2008 Asset Management Improvement Plan Based on Gap Assessment Results

Use of Information Technology in Asset Management for Sewage Treatment Plants in the Drainage Services Department

T.K. Wong

Abstract Over the past few decades, the advent of the digital revolution has transformed the landscapes of all utilities. Sewage treatment is one of them. Information technology (IT) has increasingly been incorporated into all operational aspects of the sewage treatment industry, and particularly in the last decade, it has become an indispensable tool without which the industry cannot keep going, thanks to increasing automation of sewage treatment plants. Not only is IT used exclusively for automation. An area that is receiving increasing attention lately in terms of the value added of applying IT to support business decision making and operation is asset management. In order to effectively and efficiently manage assets in sewage treatment plants, the Drainage Services Department has transformed their life-cycle management system of sewage treatment asset from paper-based processes to an IT-based system. In the heart of this IT system is an application called "Sewage Treatment Operation and Maintenance Management Information System (STOMMIS)", which involves the synchronization and direct integration of multiple IT systems, including process control and automation systems, maintenance management and stock control systems, financial information system, and laboratory information systems. Data collected by STOMMIS are stored in a data warehouse for analysis to support optimization of maintenance activities as well as the development of asset replacement plan based on asset conditions and risks. Over time, the use of information technology to support asset management will only evolve and increase. This will include the use of intelligent devices on assets that will be able to monitor and report back their capacity, use, and downtimes. This paper will give an overview on what the Drainage Services Department, the Government of the Hong Kong Special Administrative Region has done in terms of applying IT to asset management in sewage treatment plants and the way forward for continual improvement.

T.K. Wong (✉)
Drainage Services Department, The Government of the Hong Kong Special Administrative Region, Hong Kong, China
e-mail: tszkinwong@dsd.gov.hk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_79

Keywords Asset management · Information technology · Computerized maintenance management system · Enterprise asset management · Risk management

1 Introduction

Drainage Services Department (DSD) is currently operating 292 sewage treatment facilities, including 68 sewage treatment works and 224 sewage pumping stations. DSD is also managing an extensive sewerage network that covers the entire Hong Kong, with a total length of about 1,683 km, almost the distance from Hong Kong to Jinan, Shandong. Every day, on average, roughly 2.7 million m^3 of sewage was collected from the public sewerage network and treated by these facilities. The number of Hong Kong people being served by DSD's sewerage network and facilities is almost 6.7 million, about 93 % of the total population. [1].

DSD is committed to maintaining and improving the sewage collection and treatment facilities to ensure their continual efficient and effective operation. To achieve this end, DSD envisions the introduction of a "Total Asset Management Scheme" to ensure a clearer and more systematic allocation of resources [2]. Specifically, DSD is implementing PAS 55 [3], a Publicly Available Specification published by the British Standards Institution for the optimised management of physical assets, and is actively seeking certification of it. PAS 55 is chosen because this specification is the culmination of the latest thinking in terms of best practices in asset management systems. Furthermore, the International Standards Organisation (ISO) has accepted PAS 55 as the basis for development of the new ISO 55000 series of international standards.

As the challenges posed by the sewage collection network and the sewage treatment facilities are different from the perspectives of asset management and risk complexity [4], DSD has separate information systems for each. In this paper the information system for the management of DSD's sewage treatment facilities (and also 33 stormwater pumping stations, which constitute only a trivial part of the system), Sewage Treatment Operation and Maintenance Management System (STOMMIS), will be discussed.

2 Asset Management and PAS 55

PAS 55 defines asset management as "systematic and coordinated activities and practices through which an organization optimally and sustainably manages its assets and asset systems, their associated performance, risks and expenditures over their life cycles for the purpose of achieving its organizational strategic plan" [3]. It is much more than just the maintenance or care of physical assets. Good asset management considers and optimizes the conflicting priorities of asset utilization

and asset care, of short-term performance opportunities and long-term sustainability, and between capital investments and subsequent operating costs, risks and performance. "Life cycle" asset management is also more than simply the consideration of capital costs and operating costs over pre-determined asset "life" assumptions. Truly optimized, whole life asset management includes risk exposures and performance attributes, and considers the asset's economic life as the result of an optimization process (depending upon the design, utilization, maintenance, obsolescence and other factors) [5].

Overall, using PAS 55 enables DSD to [6]:

- Achieve asset management good practices
- Start processes to map the entire asset base and create the information strategy in accordance with the overall strategy
- Organize around true life cycle asset management processes
- Challenge and reduce current time-based work and replace with a "risk-based" management approach
- Position asset management-specific accountability from the "shop floor to the top floor" and create motivational performance management
- Focus on building the asset management knowledge base
- Understand and target the tools, and engage the entire organization
- Adopt a truly holistic approach by continuously challenging good or best practices.

3 Role of Information Technology in PAS 55

Unlike the 2004 version, which explicitly mandates the establishment and maintenance of an "asset management information system" [7], the role of information technology (IT) in PAS 55-1:2008 is relatively agnostic. However, it is a no-brainer that an IT system is needed for compliance, given the emphasis the 2008 version places on information management.

As a specification document that provides the "requirements" for good asset management, PAS 55 does not provide any "instructions" for selecting information systems. However, there are aspects of PAS-55 that place specific requirements upon any information systems that may be used. Most of the more direct implications for IT systems are contained in Clause 4.4.6, which requires:

- Identification of the required asset management information considering all phases of the asset life cycle
- Design, implementation and maintenance of a system(s) for managing asset management information

- Employees and other stakeholders, including contracted service providers, shall have access to the information relevant to their asset management activities or responsibilities
- Where separate asset management information systems exist, the organization shall ensure that the information provided by these systems is consistent
- Establishment, implementation and maintenance of procedure(s) for controlling all required information

So while PAS 55 does not directly mandate a particular set of IT solutions be implemented, it does set goals for IT governance and control that are best met by an application capable of encompassing the entire asset life cycle, which includes the processes of planning and engineering of the asset, maintenance and operation and the eventual retirement and decommissioning. Across this entire asset life cycle, the IT system is used to define enterprise risks and how they will be managed, which include physical failures, operational requirements, environmental events, etc. Such information must be made available to all stakeholders. This suggests that direct implementation of contemporary Computerised Maintenance Management System (CMMS) alone may not be the optimal solution. In fact, the most desirable situation is the complete integration of into a single enterprise IT system [8].

4 Configuration of STOMMIS

Before the implementation of STOMMIS, DSD had a myriad of information systems (a total of five computerized maintenance management systems, more than 10 stock or inventory management systems, a financial information system, and many "excel-based" systems, etc., see (Fig. 1), each having its own database structure and format, for the management of its sewage treatment facilities. In theory, together these systems can fulfil the analysis and reporting requirements demanded by good asset management principles, but in practice the cost of manually collating and reporting the data is not acceptable in terms of manpower requirement and in light of the availability of contemporary information systems.

To address this problem and to follow the global trend of IT systems consolidation to support decision making and operation in asset management [9], DSD has synchronized and integrate those information systems that are essential to the asset management of sewage treatment facilities and adoption of PAS 55. Hence, STOMMIS is born.

STOMMIS is an Enterprise Resource Planning (ERP) system consisting of a data warehouse with Business Intelligence (BI) capability, a CMMS complete with procurement and inventory control functions, and interfaces with various sewage treatment facilities' process control systems, Government Financial Management Information System, and Laboratory Information Management System. Resilience is provided through disaster recovery servers located at a different building from the production servers.

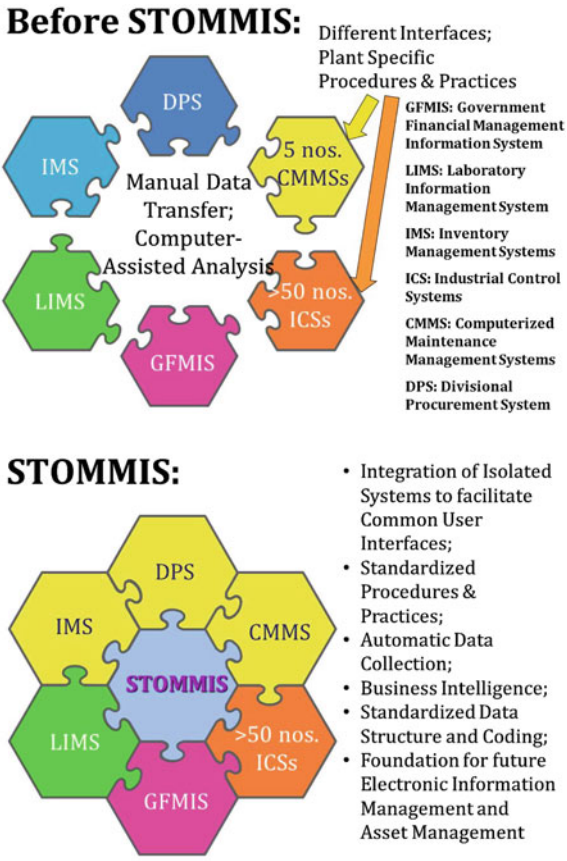


Fig. 1 Situations before and after implementation of STOMMIS

5 Role of STOMMIS in Asset Management

By integrating the various information systems, STOMMIS is able to provide the following features that support PAS 55 implementation:

- A single enterprise system encompassing the whole life cycle of asset "from cradle to grave"
- A single version of the truth stored in the data warehouse for all asset information throughout the whole DSD
- An inventory register with a well-defined hierarchy
- On-going, near real time asset performance information collected from process control systems
- A single platform for defining performance criteria and associated key performance indictor (KPI) measures for comparison across facilities both within and without DSD for continual improvement purposes

- A single, unified interface for operation and maintenance staff to run their day-to-day activities
- A single, unified interface for management staff to track and manage problems and changes
- Identification and management of asset-related risks from information collected through CMMS and process control systems
- Built-in formal structure for reporting asset failures by individual asset and by asset type, thereby supporting mean time between failures and mean time to repair types of analysis with a view to further enhancement for failure mode and effects analysis to support Reliability-Centered Maintenance (RCM)
- Workflow engine of CMMS enables procedures and practices to be standardised for the operation and maintenance of all sewage treatment facilities
- A secure data repository with disaster recovery capability for all asset information

6 Challenges and Way Forward

Asset information in STOMMIS is important for the successful implementation of asset management because it represents the collective knowledge used to manage assets. However, there are two main challenges: How the required information can be captured and what actions can be taken with this information on hand.

Information emerging from the life cycle of asset is usually hard to capture and in most cases lost. Although performance data can be automatically captured from process control systems, asset conditions cannot be easily done. The emergence of automated identification and capture (ADIC) technologies such as Radio Frequency IDentification (RFID) and smart sensors together with wireless communication can greatly help in information collection owing to their ability to capture and manage information regarding key events along an asset's life cycle [10]. They are recognised as the key constituents of eMaintenance, the technological framework that empowers organisations to streamline their asset management services and data delivery across the maintenance operations chain.

eMaintenance can be considered a technology where information is provided where it is needed and maintenance is a task that is about information when done in an effective way [11]. Maintenance actions are taken at optimal timing based on need and risks. It is a hot research topic in recent years. The EU FP6 funded Dynamite project (Dynamic Decisions in Maintenance, IP017498) developed and tested a set of methodologies and tools to support the eMaintenance processes [12]. Besides RFID and smart sensors, eMaintenance requires technologies such as signal analysis, smart decision support, portable computing devices, cloud services, common database schemas, etc. It has been argued that such technological advancements are likely to provide a bonanza in asset management [13], yet this has yet to be materialized and its economic case yet to be vindicated. eMaintenance

is still at its inchoate stage but its adoption will be more widespread with improved availability of data from RFID and smart sensors that can support diagnosis and prognosis and breakthroughs in signal analysis techniques and simulation models.

7 Conclusion

Asset management has become an essential business process in sewage treatment industry. Using PAS 55, a standard methodology for asset management, an organization can drive down costs and bring about service improvement. Information technology forms a core part of asset management, without which the whole edifice cannot stand.

STOMMIS is a case in point. It forms an indispensable part of DSD's asset management system. Although in its present form it can already satisfy the state-of-the-art good asset management practice, namely PAS 55, there is still scope for improvement. The use of RFID, smart sensors and wireless communication technologies can greatly improve the data collection process, and the use of advanced algorithms for data analysis can provide insights for decision making. However, it will take some time to prioritise all the offered opportunities and reap the full benefits.

It should be emphasized that the application of information technology to asset management is not about having one single initiative and several projects, but about creating a way of life in daily operation and maintenance activities that encourages timely capture of relevant data from smart devices, analysis of that data by improved techniques and increased computing power and their effective use in decisions and execution.

References

1. Drainage Services Department, Government of the Hong Kong Administrative Region. (2013). Drainage Services Department in Brief. Retrieved from http://www.dsd.gov.hk/EN/Files/publications_publicity/publicity_materials/leaflets_booklets_factsheets/DSD_in_Brief_English_for_web.pdf
2. Drainage Services Department, Government of the Hong Kong Administrative Region. (2012). DSD Annual Report 2010-11. Retrieved from http://www.dsd.gov.hk/EN/Files/annual_reports/1011/director.html
3. Institute of Asset Management (2008) BS PAS 55-1:2008. Asset management. Specification for the optimized management of physical assets. London, BSI
4. Fane S, Willetts J, Abeysuriya K et al (2004) Evaluating reliability and life-cycle cost for decentralized wastewater within the context of asset management. Paper presented at 1st international conference on onsite wastewater treatment and recycling (NOWRA/NOSSIG/OnSiteNZ)/6th specialist conference on small water and wastewater systems (IWA/AWA), Fremantle, Australia, 11–13 February 2004

5. Institute of Asset Management (2008) BS PAS 55-2:2008. Asset management. Guidelines for the application of PAS 55-1. BSI, London
6. IBM. (2011). IBM Software White Paper. IBM enables a natural alignment with PAS 55. Retrieval from http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&htmlfid=TIW14035USEN
7. Institute of Asset Management (2003) BS PAS 55-1:2004. Asset management. Specification for the optimized management of physical infrastructure assets. BSI, London
8. Vanneste SG, Van Wassenhove LN (1995) An integrated and structured approach to improve maintenance. Eur J Oper Res
9. Information Technology in Water and Wastewater Utilities Task Force of the Water Environment Federation (2011) Information technology in water and wastewater utilities. WEF manual of practice no. 33. WEF Press, Alexandria
10. PROMISE (2006) PROduct lifecycle management and information tracking using smart embedded systems (PROMISE). Retrieved from http://www.promise.no/
11. Gilabert E, Jantunen E, Emmanouilidis C et al (2011) Engineering asset management 2011. Proceedings of the sixth world congress on engineering asset management. Optimizing e-maintenance through intelligent data processing systems. Springer, Berlin
12. Holmberg K, Adgar A, Arnaiz A, Jantunen E, Mascolo J, Mekid S (2010) E-maintenance. Springer, Berlin
13. Jantunen E, Gilabert E, Emmanouilidis C et al (2009) e-Maintenance: a means to high overall efficiency. In: Proceedings of 4th world congress on engineering asset management (WCEAM). Springer, Berlin
14. López Campos MA, Márquez AC (2010) Modelling a maintenance management framework based on PAS 55 Standard. Qual Reliab Eng Int

Development of a Total Asset Management Strategy for the Operations and Maintenance Branch of the Drainage Services Department, the Government of the Hong Kong Special Administrative Region

Ian Martin, Edward Poon, Yiu Wing Chung, Kwai Cheung Lai and Chi Leung Wong

Abstract Established in 1989, the Drainage Services Department (DSD) of the Government of the Hong Kong Special Administrative Region (HKSAR) manages the drainage and sewerage infrastructure in Hong Kong. DSD's vision is to provide world-class wastewater and stormwater drainage services enabling the sustainable development of Hong Kong. Within this Department, the Operations and Maintenance (O&M) Branch is the arm responsible for the operation and maintenance of public stormwater drainage and sewerage assets. Being a relatively wealthy city, Hong Kong has one of the most reliable drainage and sewerage systems around the world. Being a responsible organisation, DSD's O&M Branch intends to implement Total Asset Management (TAM) to optimise the long-term management of assets and ensure cost effectiveness in utilising funds. AECOM Asia Co. Ltd. was commissioned by DSD in November 2012 to provide the consultancy support to the TAM Project in developing the TAM strategy in O&M Branch. This paper will present (1) the needs and drivers for DSD's O&M Branch in adopting TAM; (2) the

I. Martin (✉) · E. Poon
AECOM, CA, USA
e-mail: ian.martin@aecom.com

E. Poon
e-mail: edward.poon@aecom.com

Y.W. Chung · K.C. Lai · C.L. Wong
Mainland South Division, Drainage Services Department,
The Government of the HKSAR, Hong Kong, China
e-mail: yiuwingchung@dsd.gov.hk

K.C. Lai
e-mail: kclai@dsd.gov.hk

C.L. Wong
e-mail: sclwong@dsd.gov.hk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_80

approach adopted by DSD's O&M Branch in commencing its Asset Management (AM) journey; and (3) the way forward for DSD's O&M Branch in implementing TAM. This paper will also address how the O&M Branch is prepared to work together with the other DSD's Branches in expanding the current TAM strategy across the entire Department.

1 Introduction

The Drainage Services Department (DSD) of the Government of the Hong Kong Special Administrative Region (HKSAR) manages stormwater runoff and wastewater collection and disposal from a population of around seven million people. These activities are absolutely fundamental for the sustainability, resilience and well-being of a developed community. They support the fabric of a modern society and support economic growth.

These activities are enabled through an extensive network of intakes, pipes, culverts, open channels, pumping stations, treatment facilities, outlets, and many other infrastructure assets. To provide an understanding of scale, the network includes 1,647 km of sewers, 2,372 km of drains, 338 km of engineered channels, more than 160,000 nos. of manholes, 320 pumping stations and 68 wastewater treatment facilities.

Although DSD has successfully managed these networks since its inception, it has committed to deliberately progressing its systematic asset management journey.

2 Organisational Drivers

Infrastructure, such as that enabling the services that DSD delivers, has been developed progressively for more than 100 years and represents major societal investment. In many "mature" nations, it was not uncommon as recently as ten years ago, to find organisations owning and managing extensive networks unable to report what infrastructure they had, let alone where it was, what condition it was in, its value or its performance, and what level of service was being provided.

Typically, prominent drivers for organisations to implement systematic asset management practices are the following:

- Limited financial, or other resources
- Legislative or regulatory requirement
- Unknown or high risk profile
- Industry influence

However, there is a growing industry understanding of the benefits realised through implementing an appropriate level of asset management practice. Indeed, there is a growing belief that true value from implementing systematic asset

management principles is only achieved through strong organisational commitment, and that, in many ways, prescriptive legislation can lead to the adoption of a compliance mentality which can stymie innovation and hinder asset management advancement.

DSD has recognised the benefits that systematic asset management can bring to itself and its stakeholders, and is committed to achieving a level of maturity commensurate with the population served, stakeholders' needs, organisational profile, and extent and characteristics of the enabling infrastructure. Although it has, to date, had adequate funding to react to maintenance and renewal needs as they have arisen and to undertake its infrastructure development projects, it is driven by its desire, as a responsible entity, to:

- Most cost effectively provide its services to its current and future customers and to meet the needs of its stakeholders.
- Improve its customer-centric approach through identification of the stakeholders' expectations and effectively mobilisation of resources to provide the desired level of services.
- Identify and manage risk proactively by avoiding and minimising disturbance to service and social activities due to asset failure.
- Demonstrate prudent and sustainable management to its stakeholders including the general public, DSD staff and the natural environment.

There is also the basic understanding that, whilst there has been adequate funding to date, this does not necessarily mean that this will continue into the future, as many nations have experienced through the recent Global Financial Crisis.

3 Asset Management Maturity

Prior to commencing the Project, DSD undertook industry research to gain a better understanding of the path of asset maturity development in other nations, and to apply any lessons learned to its own journey. The strongest learning points were identified as:

- The need to understand that asset management is more than just managing assets. The term asset management is somewhat self-limiting and there needs to be a strong understanding that assets only exist to provide a service. Asset management describes the activities and underpinning philosophies behind successfully delivering asset intensive organisational objectives.
- The need for strong corporate commitment to asset management, and the understanding that, for asset intensive organisations, i.e. what they do to deliver the service to the customers.
- The need to establish a sound foundation on which to build processes, tools and systems. Although New Zealand and Australia have been practicing systematic asset management for nearly twenty years, it is only relatively recently that asset

management policies, strategies, objectives, and even organisationally-consistent service level, risk and criticality frameworks are being developed. This has led, in some cases, to significant rework to rationalise existing tools and frameworks to achieve a clear "line-of-sight" from organisational objectives to works programmes.
- The importance of good data, and the understanding that this is not necessarily the same thing as comprehensive data. Capturing and managing data, particularly associated with extensive underground networks, is expensive and time consuming. Data need to be adequate to enable sound decision-making, but this needs to be balanced with the effort to collect and manage it.
- The importance in good processes and systems, but also the understanding that an asset management information system is only a tool, and in itself is not the answer to sound asset management planning.

Briefing sessions coupled with pre-workshop preparatory information packs were held with all relevant O&M staff members to raise awareness of asset management principles, introduce the Project, and prepare them for the gap analysis interviews introduced in the following section. Noting the reliance on senior management commitment to achieve asset management planning success, a specific briefing was delivered to the senior directorates.

4 Developing the Total Asset Management Roadmap

Approach adopted for this Project is to conduct a gap analysis to compare existing practices against leading asset management practices from around the world and to develop a summary of gaps and the corresponding area of improvements. Each improvement will be assessed on their relative importance for practical improvement in current drainage asset management practices and then prioritised to form the Total Asset Management roadmap for the O&M Branch of DSD.

4.1 Objectives

The objectives of this Project were to:

- raise asset management awareness within the O&M Branch of DSD;
- systematically assess and document DSD's current asset management practices against a series of leading asset management practice statements sourced from a series of internationally recognised frameworks; and
- develop a prioritised, scheduled and resourced roadmap of improvement activities to guide DSD in its systematic implementation of asset management planning in O&M Branch to achieve an "appropriate" level of asset management maturity.

4.2 Assessment Framework

Frameworks used for assessing asset management practice in mature asset management nations were reviewed and from these a composite framework was developed specifically for the O&M Branch of DSD. Components that fed into this specific framework were:

- *AECOM "Gap Tool"*. This tool has been progressively developed over the past 15 years with an underpinning framework drawing from principles within the International Infrastructure Management Manual (IIMM), amongst others.
- *International Infrastructure Management Manual (IIMM).* First published in 1996, with updates in 2002, 2006 and 2011. This may be the earliest comprehensive AM manual and is internationally recognised as the leading asset management guideline.
- *BSI PAS 55 2008*. This document provides the "what you need to do" to accompany the "how you should do it" provided in the IIMM. ISO 55000 is currently being developed for release in 2013 which will supersede PAS 55.
- *WSAA Asset Management Process Benchmarking Framework.* This is the leading AM process benchmarking project across the world for the water industry and has been in place since 2004.
- *Asset Management BC Roadmap.* AssetSMART is a framework that local governments in British Columbia (Canada) use to assess their capacity to manage their assets.
- *IWA's publication "Leading Practices for Strategic Asset Management"*. In response to the need identified by its utility members, the Water Environment Research Foundation funded a research programme on Strategic Asset Management (SAM) Implementation and Communication for wastewater and water utilities. The document identifies and documents leading strategic asset management practices used by utilities

The composite framework has 360 "leading practice statements" which systematically guide the assessment of the O&M Branch of DSD's:

- Asset knowledge, the appropriateness, reliability and accessibility of data, and the processes associated with the use and maintenance of asset data.
- Strategic planning processes, the processes used in the implementation of AM activities including failure planning, risk management, service level reviews, and long term financial planning.
- Current AM practices, the processes used in the implementation of AM activities including capital expenditure programmes, and operations and maintenance management.
- AM plans that identify the optimum lifecycle management tactics and resources.
- Information systems to support (and often replicate) AM processes and store/manipulate data.
- Organisational tactics including organisational, contractual and people issues.

- Embedment of AM into the organisation through formal processes and organisational structure and commitment (PAS 55).

Each leading practice statement is weighted to reflect its relative importance within its corresponding element. Each element has been assumed to be weighted evenly. These are user-defined weightings which can be reviewed and refined at subsequent assessments if required (Fig. 1).

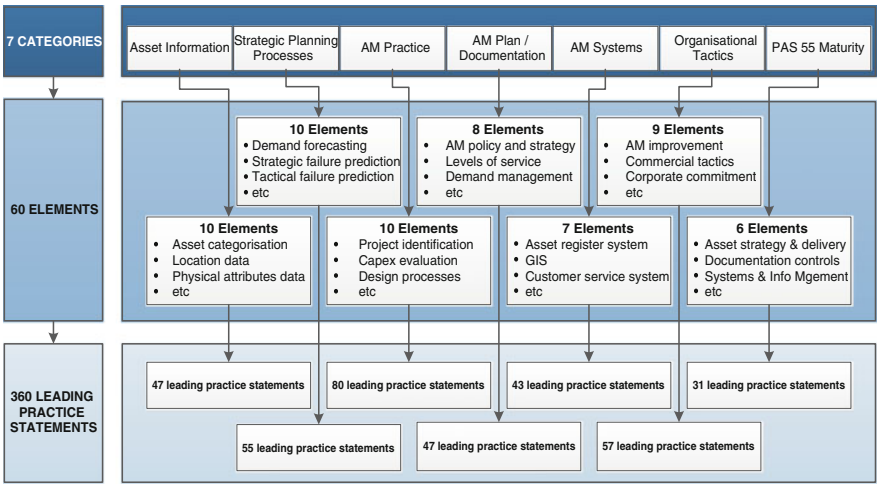


Fig. 1 DSD's asset management practice assessment framework

4.3 Assessment of Asset Management Practice

A two-week series of workshops was held with key staff in O&M Branch to assess current and "appropriate" practice, guided by the developed framework. Each current practice was described and scored with reference to the defined scoring guidelines presented in Table 1. This was then compared to "appropriate" practice, and a description of what would be required to achieve this appropriate practice note. Current improvement initiatives were also recorded.

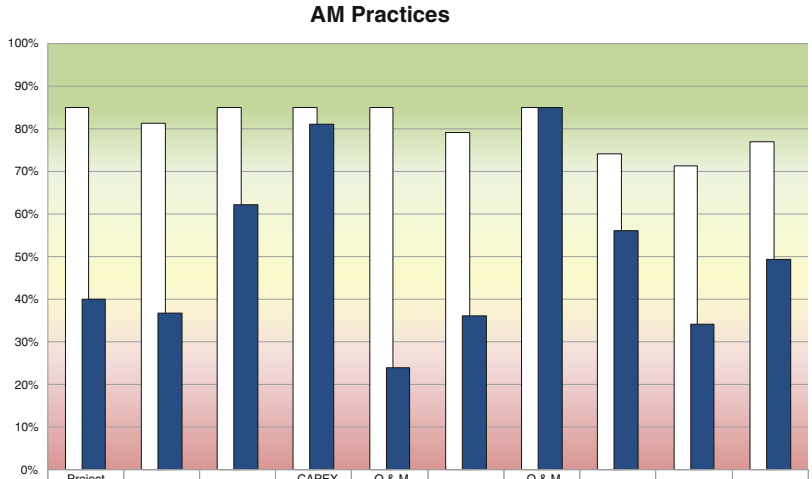
Appropriate practice needs to be defined by the organisation itself, and typically draws on an understanding of the organisation's risk profile, its objectives and priorities, the size and complexity of the service and enabling infrastructure it manages, how it compares with other peer organisations (e.g. through benchmarking programmes) and its available resources.

This can be difficult to define in the initial status review, therefore AM specialists provided guidance based on their industry knowledge and targeted those improvement actions to be implemented within a suitable time frame. Subsequent reviews will allow organisations within the O&M Branch of DSD to progressively define their own "appropriate" practice.

Asset management practice "maturity" was presented both graphically and in tabular format to allow the DSD's management to quickly understand the magnitude of the gap between current and appropriate practice, and also be able to drill down into the detailed activities required to close this gap. An example of the graphical presentations is provided below (Fig. 2).

Table 1 Composite benefit

Normalised Gap	Assessed Benefit		
	High	Moderate	Low
100	Very High	High	Moderate
90	Very High	High	Moderate
80	Very High	High	Moderate
70	Very High	High	Moderate
60	High	Moderate	Low
50	High	Moderate	Low
40	High	Moderate	Low
30	High	Moderate	Low
20	High	Moderate	Low
10	Moderate	Moderate	Low
0	Moderate	Low	Low



	Project Identification/ Prioritisation	CAPEX Evaluation	Design Processes	CAPEX Contract Management	O & M Strategy and Analysis	O & M Plans	O & M Contract Management	Operations Execution	Maintenance Execution	O & M Monitoring
	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	3.10
□Appropriate Score	85%	81%	85%	85%	85%	79%	85%	74%	71%	77%
■Current Score	40%	37%	62%	81%	24%	36%	85%	56%	34%	49%

Fig. 2 Example graphical outputs of AM practice "maturity"

Description	Process	Information Systems	Asset Knowledge (Data and plans)	Leadership	Human Resources
Innocence	No process exists. Never do this.	No system exists	No results seen. No confidence in information. Planning based on very large unsupported assumptions.	Senior management has yet to develop and launch quality improvement as a major strategic goal. There is no real commitment by leadership on making quality a high priority. Senior managers are not engaged with customers, suppliers, employees, and others on how to improve quality. There is no concern for things like social responsibility.	Training and fundamental development of employees is not widely practiced. Few employees are empowered and no priority is given to building the human resources of the company. Reward and recognition programs are not focused on employee performance and quality improvement. A few managers support human resource development, but no real leadership exists from the top.
Awareness	Minimal documentation. Ad hoc procedures. Occasionally do this.	Manual system exists or plans for automated systems are in place. Some very basic user needs met.	Minimal results, long way to go. Very low data confidence.	A few senior level executives are supportive of quality improvement. Employees are encouraged to become more involved. Communication is top down and not across organizational boundaries. Continuous improvement programs are up and running in some parts of the company. Overall organizational policies do support quality improvement.	Employee empowerment is not encouraged throughout the company. Rewards and recognition for quality improvement is not fully deployed at all levels within the company. Most of the focus is on individual employee recognition and not teams or groups. Not all employee development programs are linked to the company's strategy and quality objectives. The organization does not consistently support the needs and requirements of
Systematic Approach	Semi formal process. Completed on an as-needed basis for critical programs and activities.	Automated system exists. Basic user needs met.	Some results, still below expectations. Low data confidence.	Senior level managers are sharing their ideas on quality improvement with customers, employees, suppliers, and other key players. Management performance is linked to quality. Many parts of the organization are actively engaged in quality programs. Senior level leadership is supportive of strategic quality initiatives.	Many parts of the organization have empowered their employees through cross functional teams and the sharing of information. The company has an overall plan to fully develop its human resource capital. Employees are rewarded for making improvements to quality. Management supports the development of employees in many parts of the organization. The organization is sensitive to the needs
Competence	Formal process exists and documented but still evolving. Often do this on many programs.	Good system in place. Widely available. All key user needs met.	Good results, getting there. Reasonable data confidence.	Most senior level managers are visibly involved in quality improvement. Senior managers are meeting with teams, suppliers, and other key players in process improvements. Management behaviour at all levels in the organization reflects a commitment to quality. Senior management is actively promoting and communicating quality improvement.	Senior management and most middle level managers are very supportive of strong human resource practices to build and develop employees. Work teams and groups are empowered, providing valuable improvements in almost every part of the business. Employees have quick access to data for analysis and sharing of information in most parts of the company. Employee ideas for making improvements is strongly encouraged and acted upon throughout most of the organization.
Excellence	Formal documented process, well tested and followed. Usually do this, omitted only in exceptional circumstances.	Strong system in place. Nearly all user needs met.	Excellent results, still some room to improve. Good level of data confidence.	Senior level management is strongly involved and behind quality improvement within the company. Management is very supportive and working to form teams throughout the company. Senior managers are communicating clearly the vision and goals behind quality improvement and how it must interact with customers, suppliers, employees, and others in the value chain. Senior management is very committed to all continuous improvement efforts.	Wide implementation and integration of employee growth plans, including training programs, career development paths, evaluation / self-awareness processes, compensation, empowerment, and measurable results. Good levels of involvement by employees in day to day operations and planning the business. People work well within teams and across organizational functions. Recognition programs are in place for rewarding employees who improve quality. The organization is aware to the needs and requirements of employees, and is working to make sure employees are productive and satisfied.
Best Possible	Strictly formal process. Always do this, standard operating procedure. Process heavily emphasised, not deviated from.	State-of-the-art system in place. All user needs met.	Unparalled results, a total success. Very high level of data confidence.	Senior level management is fully committed to quality improvement within the company as a key performance factor. Management operates as teams throughout the company. Senior managers are communicating clearly and regularly the vision and goals behind quality improvement and how it must interact with customers, suppliers, employees, and others in the value chain. Senior management is fully committed to all continuous improvement efforts.	Full implementation and integration of employee growth plans, including training programs, career development paths, evaluation / self-awareness processes, compensation, empowerment, and measurable results. Very high levels of involvement by employees in day to day operations and planning the business. People work well within teams and across organizational functions. Strong recognition programs are in place for rewarding employees who improve quality. The organization is very sensitive to the needs and requirements of employees, working hard to make sure employees are productive and satisfied.

The current status of the O&M Branch of DSD is typical for an infrastructure intensive organisation commencing its systematic asset management development programme. Areas that are typically of interest to engineers and other technical staff include data and systems, and works control. Areas of strength were identified as:

- Staff understand the components of asset management and appear to work collaboratively, even though this Project marks the start of DSD's systematic asset management development journey and much of the asset management terminology is new to the DSD.
- Sound procurement and contract monitoring practice is in place with clear guidelines and procedures developed and followed.
- Hydraulic modelling is undertaken by DSD. Although not necessarily covering every pipelines due to limitation on resources and software, these models cover most of the wider Hong Kong area.
- There are extensive and frequent condition data capture programmes in place, mainly using CCTV techniques, but also zoom camera, sonar and laser profilometry.

Key opportunities for improvement were identified as:

- An asset management policy, strategy and objectives, together with an integrated asset management framework are needed as foundation for sustainable management of the infrastructure and the service enabled by this infrastructure.
- Service level, performance and risk management frameworks are required as a basis on which to define what DSD should provide to the customers and other stakeholders in respect of operations and maintenance of sewerage and drainage assets, which in turn forms the basis for all works activities and expenditure.
- Documented lifecycle strategies and an asset management plan are needed to formally bring together the different aspects of asset management planning and to communicate current and future issues, strategies to address these issues, work requirements and expenditure forecasts. The asset management plan should be seen as the DSD's main planning tool.

4.4 Developing the Roadmap

In order to develop a clear and achievable roadmap to guide the O&M Branch of DSD for the necessary improvements to be carried out in short-medium term say the next three-year, outputs from the gap assessment were then grouped, prioritised and scheduled, with responsibilities and estimated resource requirements recorded, and detailed project briefs prepared for each improvement project scheduled to commence in the next 12 months.

Individual improvement activities were aggregated into improvement projects based on the areas of improvement which may be to address the needs of one or more than one elements and/or categories of the framework, which had, in some

cases, sub-projects. Each improvement project was prioritised and scheduled considering the following aspects:

4.4.1 Benefit

Composite benefit has been assessed as Low, Moderate, High or Very High considering:

- Size of the weighted gap within the gap analysis, which assesses both the size of the gap between current and appropriate practice at the elemental level and the assessed importance of that particular practice. This has been "normalised" to achieve a range of scores between 0 and 100.
- Assessment of the benefit of implementing the identified improvement project. This has been rated as Low, Moderate and High.

4.4.2 Effort

Composite effort has been assessed as Low, Moderate, High or Very High considering:

- *Estimated cost of implementing the improvement project*. This is a high level estimate only and will be reviewed in greater detail prior to commissioning external resources and/or otherwise implementing the improvement project. Costs have been rated as < $100,000, $100,000–$500,000, $500,000–$1 million, and > $1 million.
- *Complexity and/or risk associated with undertaking the improvement project*. This has been rated as Complex, Moderate and Straightforward (Table 2).

4.4.3 Priority

Priority then considers the composite benefit and composite effort as presented in Table 3.

Table 2 Composite effort

		Complexity/Risk		
		Complex	**Moderate**	**Straightforward**
Estimated Cost	**>$1,000,000**	Very High	Very High	High
	$500,000 - $1,000,000	Very High	High	Moderate
	$100,000 - $500,000	High	Moderate	Low
	<$100,000	Moderate	Low	Low

Table 3 Improvement project priority

Composite Benefit	Composite Effort: Low	Moderate	High	Very High
Very High	Very High	Very High	High	Moderate
High	Very High	High	High	Moderate
Moderate	High	Moderate	Moderate	Low
Low	Moderate	Low	Low	Low

4.4.4 Scheduling

Scheduling of improvement projects was undertaken by considering the individual improvement activities and the logical progression of these. The fundamental principles that were followed in developing the improvement plan were defined through the initial work reviewing the lessons learned from others in asset management development discussed. One of the key principles is that the development of a sound policy and framework is essential for achieving a clear line of sight, "joined-up" thinking, consistent and robust asset management processes, systems, strategies, and plans.

The scheduling recognised instances where completion of improvement projects relies on the completion of activities in other improvement projects. This has resulted in some cases in Moderate priority improvement projects being scheduled before High priority improvement projects.

Figure 3 and Table 4 present the improvement programme. Each improvement project is detailed into improvement activities, not presented in Table 4 for clarity, and, underpinning this information is a series of detailed project briefs.

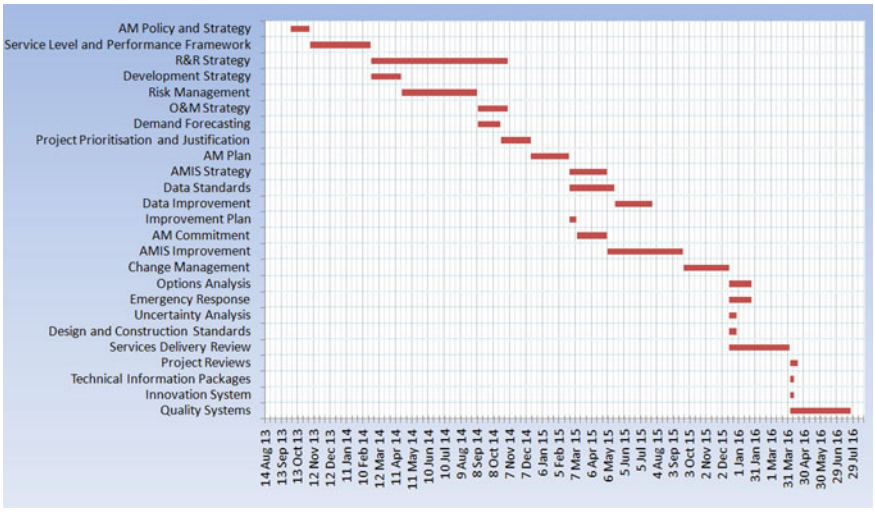


Fig. 3 Draft improvement programme

Table 4 Improvement projects and underpinning improvement activities

Improvement Project		Sub-Project	Composite Benefit	Composite Effort	Priority
1	AM Policy and Strategy	AM system	H	L	VH
		AM policy and strategy			
		Review			
2	Service Level and Performance Framework	Performance Framework	H	H	H
		LOS Options and Change			
		LOS Outputs			
	R&R Strategy ***(underway)***	Lifecycle Strategies	H	VH	M
		Analysis and ODM			
4	Development Strategy	Lifecycle Strategies	H	M	H
		ODM			
5	Risk Management	Policy Strategy and Framework	H	H	H
		Risk Assessment			
		ODM			
6	O&M Strategy	Lifecycle Strategies	H	H	H
		Analysis			
		Procedures			
7	Demand Forecasting	Demand Forecasting	M	M	M
		Demand Modelling			
8	Improvement Plan		M	L	H
9	AM Plan	AM plan	VH	M	VH
10	AM Commitment	Responsibilities	H	H	H
		Organisational structure			
		Training and skills			
11	Uncertainty Analysis		M	M	H
12	AMIS Strategy	AMIS Strategy	H	M	H
		Asset Register System			
		Geographic Information System			
		Customer Service System			
		Maintenance Management System			
		Condition Monitoring System			
		Advanced AM Systems			

(continued)

Table 4 (continued)

Improvement Project		Sub-Project	Composite Benefit	Composite Effort	Priority
13	Data Standards	Asset Categorisation	H	M	H
		Location Data			
		Condition Data			
		Capacity Data			
		Financial Data			
		Asset Register System			
		Asset Register System			
		Geographic Information System			
		Data warehouse			
14	Data Improvement	Location Data	H	M	H
		Physical Attributes Data			
		Capacity Data			
		Asset Life Data			
		Financial Data			
		O & M Monitoring			
		Capacity / Utilisation (Hydraulic, Plant) Models			
15	AMIS Improvement	O & M Data	H	M	H
		Asset Life Data			
		Financial Data			
		Asset Register System			
		Geographic Information System			
		Customer Service System			
		Maintenance Management System			
		Condition Monitoring System			
		Advanced AM Systems			
16	Change Management	Location Data	M	M	M
		O & M Data			
		Condition Data			
		Capacity Data			
		Performance Data			
		Asset Life Data			

(continued)

Table 4 (continued)

Improvement Project		Sub-Project	Composite Benefit	Composite Effort	Priority
		Financial Data			
		O & M Monitoring			
		Geographic Information System			
		Customer Service System			
		Maintenance Management System			
		Maintenance Management System			
		External drivers			
17	Options Analysis		M	H	M
18	Project Prioritisation and Justification		M	H	M
19	Emergency Response		M	M	M
20	Design and Construction Standards		L	L	M
21	Service Delivery Review	Service Delivery Strategy	M	H	M
		Work Efficiency			
		Systems			
		Benchmarking			
22	Project Reviews		L	L	M
23	Technical Information Packages		L	L	M
24	Innovation System		L	L	M
25	Quality System		L	H	L

5 Implementing the Roadmap

At the time of writing this paper, the O&M Branch of DSD was in the process of planning the implementation of the roadmap. Key challenges identified and to be addressed include:

- *Current organisational structure*. Like most if not all organisations, DSD has several distinct Branches in charge of varying responsibilities. In this instance, the O&M Branch has undertaken the assessment and development of the roadmap, although asset management planning, in the wider definition of the term, extends beyond this Branch alone. In particular, when considering the intent of PAS 55 and ISO 55000, it could be argued that consistent frameworks and approaches should be developed across the entire DSD. Although there is reasonable communication and collaboration across Branches, there is a risk that

systematic asset management planning objectives and strategy are not aligned between Branches. Top management support and involvement is important for the successful implementation of asset management. In addition, an AM steering committee is also setup within O&M Branch where representatives from other Branches will be invited to attend the committee meetings to ensure key AM principles and directions are aligned and to enhance exchange of experience and knowledge in AM development between different Branches. For matters affecting DSD as a whole such as AM Policy and Strategy, commitment from top management will be sought to ensure approaches adopted by different Branches will be converged to the same goals.

- *Time and resource constraints*. Although asset management planning should be considered as a core activity, DSD, like most organisations at the start of the asset management journey, need to be able to manage conflicting priorities. The roadmap focuses on improvement projects of a strategic nature. It can be difficult for improvement projects to be given the priority in amongst the many day-to-day operational activities that DSD staff are faced with. Consideration is being given to separating out roles and responsibilities to address this issue.
- *Skill constraints*. DSD understands that it does not have all the expertise required to complete the improvement projects and it is likely to require external specialist assistance. However, DSD also understands that it needs to be heavily involved in the improvement process and cannot simply rely on external specialists to undertake the work for them. Asset management success will rely on organisational ownership of the process and the outputs.
- *Funding constraints and organisational priorities*. Implementing sound asset management practice, including streamlining systems and developing processes and tools will require effort and cost. Given that there have been no catastrophic asset failures and that funding is currently sufficient to respond to incidents as they arise, there is a possibility that the continued implementation of the asset management improvement roadmap is not regarded as an organisational priority.

Research on the Maturity Evaluation Method of the Transfer Phase in Flight Test

Wenjin Zhang, Jie Meng, Nan Lan and Ying Ma

Abstract In the development of new aircrafts, flight test is a crucial link between the manufacturing process of the prototype and the acceptance of the final product. Based on the division of test phrase in China, this chapter explores the construction ideas of flight test maturity models. In terms of the technology status, integration status, manufacturing status, flight test status and RMS status, this chapter describes the flight test maturity, and establishes a more comprehensive and objective model —the Flight Test Maturity Model (FTMM) with specific description of each level of the test flight maturity characteristics. In this chapter, Flight state of maturity serves as the theoretical basis for the evaluation of the flight test phase, the Attribute Comprehensive Evaluation method is also used to test flight maturity. An example is given using flight testing maturity model to evaluate the test status of a new type of aircraft, which verifies the correctness of the model and its applicability in the flight test work.

1 Introduction

Flight test refers to the flight launched to testify and examine the capability and operating characteristic of airplanes or other aerial equipment, which is also called test flight. The improvement in the reliability, maintainability, supportability and testability (RMS) of equipment is highly necessary in the development of airplanes. Flight test phase is an important phase during which the RMS of new airplanes will be evaluated. The new generation of fighter jets stands out for their new technological characteristics like supersonic cruise, stealth performance, high maneuverability and the application of comprehensive avionic technologies, which has raised new requirements for the RMS evaluation [1]. With the complexity of newly

W. Zhang · J. Meng (✉) · N. Lan · Y. Ma
School of Reliability and Systems Engineering, Beihang University, Beijing, China
e-mail: mengjie@dse.buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_81

developed fighter jets increasing, it has become quite important to search for new RMS evaluation methods [2].

The problems existing in the current evaluation methods during transfer phase of the flight test are as follows: (1) The evaluation mainly focuses on whether the capability of airplanes can meet the technical index, regardless of the influence caused by RMS. (2) The evaluation depends a lot on experience, lacking in a complete evaluation system.

Maturity Model(MM) originally comes from the research on life cycle, and is now applied in many fields, for instance, Capability Maturity Model for Software, Project Management Maturity Model and Information technology maturity model. These models have played positive roles in the guidance of engineering applications.

To better support Interdisciplinary groups, concurrent engineering and some other highly-automatic and hybrid engineering development environment, the Software Engineering Institute (SEI) in Carnegie Mellon University(CMU), USA released the Capability Maturity Model Integration (CMMI) in 2001.12 [3]. Wang put forward the concept of Reliability Engineering Capability in his master thesis, and built the framework of the Reliability Engineering Capability Maturity Model, which amplified the evaluation object [4].

This chapter combines flight test with the theory of maturity, and conducts research on a structured model that can describe and measure the maturity of flight test, which will provide certain theoretical knowledge for improvement of flight test management.

2 Flight Test Phases

The flight test of military airplanes in our country can be divided into 5 phases, including maiden flight, development flight test, evaluation flight test, verification flight test and acceptance flight test. The main object of each phase is shown in Table 1 [5].

3 The Construction of Flight Test Maturity Model

3.1 The Levels of Flight Test Maturity

Corresponding to the different phases of flight test, flight test maturity can be divided into 6 levels, as shown in Table 2 below.

Table 1 Main object of each phase

Phases	Main object
Maiden flight test	To test the aerodynamic performance, structure, maneuverability and motive power of the airplane
Development flight test	To test the flight performance and operation stability of the airplane, and to adjust its engine, subsystems and airborne equipment
Evaluation flight test	To examine the capability and technical index of the final products, verify the capability of the engine, airborne equipment and the electronic system, and evaluate the compatibility between equipment
Verification flight test	To verify whether the airplane meets the requirements raised in related documents by means of flight test
Acceptance flight test	To verify the capability and quality of the airplane and airborne equipment according to the rules for acceptance

Table 2 Matching cases between maturity model and the flight test process

Maturity level	Flight test maturity	Flight test phases of military airplanes	The development phases of weaponry
Pre-test level	FTML0	Before flight test	Development phase
Maiden test level	FTML1	Maiden flight test	
Development test level	FTML2	Development flight test	
Evaluation test level	FTML3	Evaluation flight test	Design finalization phase
Verification test level	FTML4	Verification flight test	Manufacturing finalization phase
Acceptance test level	FTML5	Acceptance flight test	Small-batch production phase

3.2 Analysis on Dimensions of Flight Test Maturity Condition

Every level of the flight test maturity is to be analyzed from different dimensions. This chapter conducts the analysis form the following five dimensions. The details of these dimensions are shown in Table 3 below.

3.3 Description of the Flight Test Maturity Model

The evaluation of Flight Test Maturity Model (FTMM) is based on the division of flight test maturity levels. According to the standard of each dimension, the flight test work can be evaluated, thus generating a stepped route for the improvement of flight test maturity. The route is shown in Fig. 1 below.

Table 3 Dimensions of flight test maturity condition

D-No.	Dimensions	Content
I_1	Technical condition dimension	Verification on the capability of certain technology
I_2	Manufacturing condition dimension	Verification on the manufacturability and economic feasibility of the technology, including the evaluation on manufacturing process, raw materials, cost, human resources and management
I_3	Integration condition dimension	Verification on the entity/physical attribute of the integration between two technical components (integration interface or integration standard), considering the mutual influence, compatibility and reliability between the two components
I_4	Examination condition dimension	Research on the flight mechanism and flight laws, verification on the capability of the airplane, engine and airborne equipment
I_5 [8]	RMS condition dimension	Verification on the reliability, maintainability, supportability and testability index of the airplane

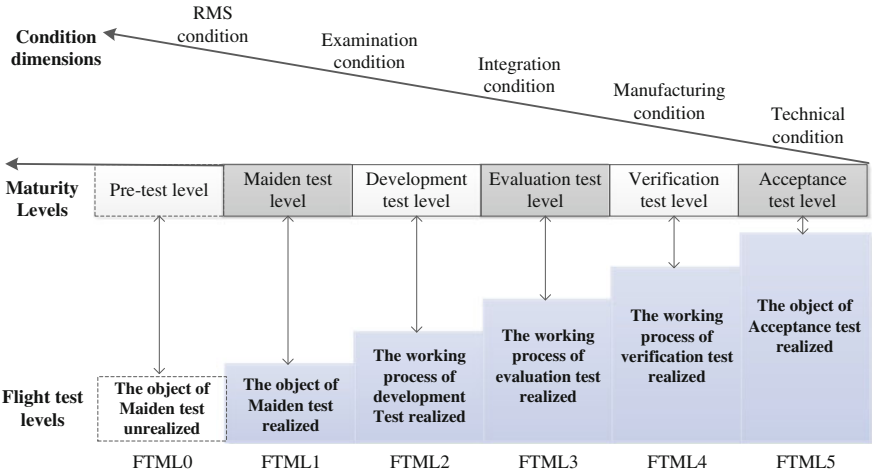


Fig. 1 The framework of FTMM

4 Research on the Comprehensive Evaluation Method of the Flight Test Transfer Phase

4.1 Basic Ideas About the Comprehensive Evaluation Method

The process about the comprehensive evaluation method is shown in Fig. 2.

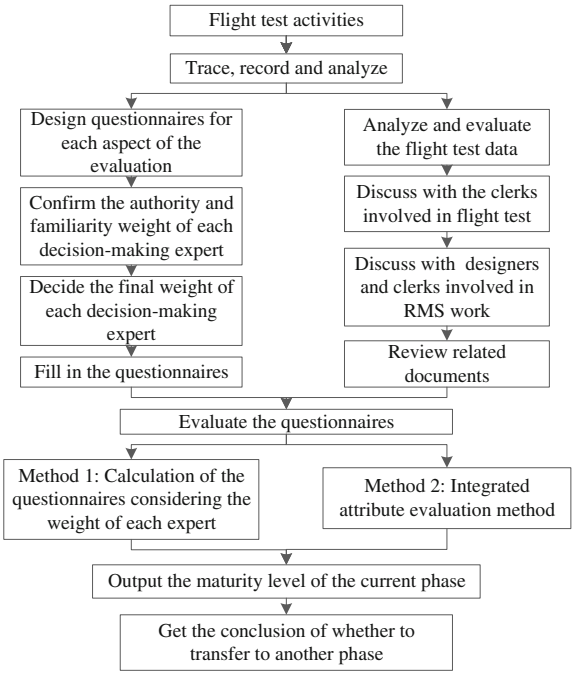


Fig. 2 The comprehensive evaluation method of the flight test transfer phase

4.2 Questionnaire Survey of Flight Test Maturity

The weight coefficient of each expert can be determined based on the following principles: the authority principle, the familiarity principle and the decision consistency principle [6].

The weight of the expert determined according to the authority principle is called authority weight. Marked as λ_i, authority weight can be judged from the qualification, popularity and academic level of the expert. The weight determined according to the familiarity principle is called familiarity weight, which is as η_i and can be judged from the expert's research area. While, the weight determined according to the decision consistency principle is called decision weight. Decision weight is marked as γ_i, a lower γ_i indicates a higher difference between decisions.

$$\gamma_i = \frac{\frac{1}{\varepsilon_i}}{\sum_{i=1}^{n} \frac{1}{\varepsilon_i}}, (\mathrm{i} = 1, 2, \ldots, \mathrm{n}), \text{ where,}$$

$$\varepsilon_i = \sum_{j=1}^{n} d_{ij} (\mathrm{i,j} = 1, 2, \ldots, \mathrm{n}), \; d_{ij} = \left[\sum_{k=1}^{m} \left(w_k^i - w_k^j \right)^2 \right]^{1/2}, (i, j = 1, 2, \ldots n)$$

In this equation, ε_i indicates the similarity between the decisions of the object expert's and other experts'. A lower ε_i indicates a higher similarity.

Therefore, the final weight of the object expert can be calculated with the following equation.

$$\lambda_{\mathrm{i}} = k_1 \lambda_i + k_2 \eta_{\mathrm{i}} + k_3 \gamma_{\mathrm{i}}$$

where k_1, k_2, k_3 represent the proportionality coefficient of authority weight, familiarity weight and decision weight respectively, and $k_1 + k_2 + k_3 = 1, 0 \leq k_i \leq 1, (i = 1, 2, 3)$.

4.3 Calculation of the Questionnaires Considering Expert Weight

The calculation method utilizes the dialogical way with the corresponding expert to calculate the number of questions that has been answered in the questionnaire, thus deciding whether certain maturity level has been reached. The detailed procedures are explained below.

(1) Determine the weight of each expert

Suppose n experts take part in the evaluation, and the weight of each decision-maker E_i is λ_i, where $0 \leq \lambda_i \leq 1$ and $\sum_{i=1}^{n} \lambda_i = 1$, $i = 1, 2 \ldots n$

(2) Fill in the questionnaires

The problems of the questionnaires are set according to the activities in each level of the maturity model. The questionnaires are filled in by the evaluation team members, in which the answers to the questions are divided into three kinds: "completely realized", "partly realized" and "not realized".

(3) Statistic Analysis of the questionnaires

The "completely realized" answer can get 1 mark, while the "partly realized" answer 0.67 marks and the "not realized" answer 0. On Level m of the flight test maturity, the number of the "completely realized" answers in Dimension j is marked as o_{mj}, while the number of the other two kinds of answers are marked as p_{mj} and q_{mj} respectively. The scores of each dimension are calculated and shown in Table 4:

Considering expert weight, the score proportion of Dimension j is:

$$\mu_j(m) = \frac{\lambda_1 a_{mj1} + \lambda_2 a_{mj2} + \cdots + \lambda_n a_{mjn}}{\lambda_1 + \lambda_2 + \cdots + \lambda_n} = \frac{\sum_{i=1}^{n} \lambda_i a_{mji}}{\sum_{i=1}^{n} \lambda_i} = \sum_{i=1}^{n} \lambda_i a_{mji}$$

where $1 \leq i \leq n$, $1 \leq j \leq 5$, $1 \leq m \leq 5$.

Table 4 Scores of each dimension

Dimension NO.	Total number of answers	Answers	Number of each kind of answer	Marks	Total score	Score proportion
I_j	N_{mj}	Completely realized	o_{mj}	1	$a_{mj} = 1 \times o_{mj} + 0.67 \times p_{mj}$	$\frac{a_{mj}}{N_{mj}}$
		Partly realized	p_{mj}	0.67		
		Not realized	q_{mj}	0		

(4) Rules for Judgment

According to the practical situation, the evaluation team can set a threshold $k_j(m)$ for $\mu_j(m)$, the score proportion of each dimension. For $j = 1, 2, 3, 4, 5$, if the inequality $\mu_j(m) \geq k_j(m)$ can be always met, the flight test maturity can be decided to have reached Level m, and it is enough for the transfer to the next phase. While if any item in $j = 1, 2, 3, 4, 5$ cannot meet the inequality $\mu_j(m) \geq k_j(m)$, based on the cask principle, the flight test maturity can be decided not to have reached Level m, in which case some improvements should be made and another maturity evaluation should be conducted.

The transfer phase evaluation is conducted following the process shown in Fig. 3.

4.4 Attribute Comprehensive Evaluation Method

(1) The Determination of Attribute Measure Based on Expert Questionnaire

The determination of attribute measure $\mu_x(A)$ is a key problem in the application of attribute mathematics, which is done according to specific problems, experiment data, expert experience and certain mathematical methods. Another approach is to use statistical methods to calculate the approximate value of $\mu_x(A)$.

Based on the definitions of attribute mathematics, the attribute space F corresponding to flight test maturity can be divided into 5 levels, which are as follows: $C_1 = \{\text{Level 1}\}$, $C_2 = \{\text{ Level 2}\}$, $C_3 = \{\text{ Level 3}\}$, $C_4 = \{\text{ Level 4}\}$, $C_5 = \{\text{ Level 5}\}$. A higher level indicates that the technologies are more mature.

(a) Statistic analysis of the evaluation questionnaires

As was talked about in the last section, for attribute space C_m, the scores of its dimensions are shown in Table 4.

(b) Determination of the attribute measure of each dimension

The score proportion of Expert i is: $a_{mji} = \frac{a_{mj}}{N_{mj}}\big|_{E_i}$

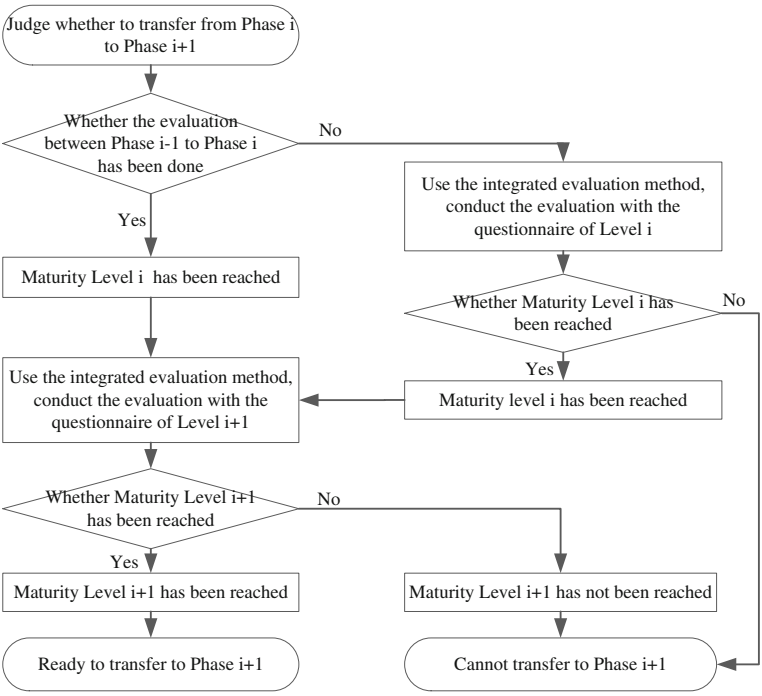


Fig. 3 The comprehensive evaluation method of the flight test transfer phase

where $1 \le i \le n, 1 \le j \le 5, 1 \le m \le 5$

Since the sum of all attribute measures is 1, the score proportion should be normalization processed.

$$a'_{mji} = \frac{a_{mji}}{\sum_{m=1}^{5} a_{mji}}, 1 \le j \le 5, 1 \le m \le 5.$$

To reflect the integrity of evaluation information, the structural dimensions are considered in the attribute measure equation.

$$\mu_x(C_m) = \frac{\lambda_1 a'_{mj1} + \lambda_2 a'_{mj2} + \cdots + \lambda_n a'_{mjn}}{\lambda_1 + \lambda_2 + \cdots + \lambda_n} = \frac{\sum_{i=1}^{n} \lambda_i a'_{mji}}{\sum_{i=1}^{n} \lambda_i} = \sum_{i=1}^{n} \lambda_i a'_{mji}$$

where $1 \le i \le n, 1 \le j \le 5, 1 \le m \le 5$

c) Proof additivity property

Following is the proof that $\mu_x(C_k)$ meets the additivity property of attribute measure [7]:

(i) $\mu_x(\mathrm{A}) \geq 0, \forall \mathrm{A} \in \mathrm{R}$;
(ii) $\mu_{\mathrm{x}}(F) = 1$;
(iii) If $Ai \in R, A_i \cap A_j = \phi(i = j)$, then $\mu_{\mathrm{x}}(\bigcup_i A_i) = \sum_i \mu_{\mathrm{x}}(A_i)$

Proof Based on the mathematical model of the comprehensive evaluation method, it can be known that $(C_1, C_2, \ldots, C_9)$ is the division of attribute space F, and $C_i \cap C_j = \phi, i \neq j$. □

Set $\Re = \{B|B = \bigcup_{i=1}^{m} A_i, A_i \in \{\phi, C_1, C_2, \ldots, C_9\}, 1 \leq n \leq 9\}$, then $\{F, \Re\}$ is the attribute measurable space.

Since $0 \leq a'_{mji} \leq 1$, and $0 < \lambda_i < 1$, then $\mu_x(C_m) = \sum_{i=1}^{n} \lambda_i a'_{mji} \geq 0$;

$$\sum_{m=1}^{9} a'_{mji} = 1$$

$$\mu_x(F) = \mu_x(C_1 \cup C_2 \ldots \cup C_9) = \lambda_1 \sum_{m=1}^{9} a'_{mj1} + \lambda_2 \sum_{m=1}^{9} a'_{mj2} + \cdots \lambda_n \sum_{m=1}^{9} a'_{mjn}$$

$$= \lambda_1 + \lambda_2 + \cdots \lambda_n = \sum_{i=1}^{n} \lambda_i = 1$$

Since $C_m \in \Re$, and $C_p \cap C_q = \phi(p \neq q)$, to make it simple, pick $C_p, C_q(p \neq q)$ randomly, then $\mu_x(C_p \cup C_q) = \sum_{i=1}^{n} \lambda_i(a'_{pji} + a'_{qji}) = \sum_{i=1}^{n} \lambda_i a'_{pji} + \sum_{i=1}^{n} \lambda_i a'_{qji} = \mu_x(C_p) + \mu_x(C_q)$.

In the same way, 3, 4 or 5, C_m can be picked randomly, where $C_m \in \Re, 1 \leq m \leq 5$, then $\mu_x(\bigcup_i C_i) = \sum_i \mu_x(C_i)$.

The proof above proves that the attribute measure equation $\mu_x(C_m) = \sum_{i=1}^{n} \lambda_i a'_{mji}$ in this chapter meets the additivity property of attribute measure.

(2) Analysis of the Comprehensive Attribute Measure of Flight Test Maturity

As was talked about in the last section, the attribute measure with single dimension $I_j(1 \leq j \leq 5)$ can be calculated through expert questionnaires. In this section, analysis of multi-dimension attribute measure will be discussed, namely, analysis of the comprehensive attribute measure of flight test maturity.

The dimension set of flight test maturity is $\{I_1, I_2, I_3, I_4, I_5\}$. According to the comprehensive evaluation method, the attribute measure classification criterion

matrix corresponding to dimension $I_j(1 \le j \le 5)$ and evaluation classes $C_1, C_2, \ldots, C_5$ is as follows.

$$R = \begin{array}{c} \begin{matrix} C_1 & \cdots & C_5 \end{matrix} \\ \begin{pmatrix} \mu_{11} & \cdots & \mu_{15} \\ \vdots & \ddots & \vdots \\ \mu_{51} & \cdots & \mu_{55} \end{pmatrix} \end{array} \begin{matrix} I_1 \\ \vdots \\ I_5 \end{matrix}$$

The comprehensive attribute measure of evaluation classes $C_1, C_2, \ldots, C_5$ is $\mu_m = \sum_{j=1}^{5} \mu_{im}, 1 \le m \le 5$.

(3) Attribute Recognition

Based on the comprehensive evaluation method, for ordered evaluation classes $(C_1, C_2, \ldots, C_5)$, the confidence criterion should be utilized to recognize which class the current flight test maturity T belongs to. Since the growth of flight test maturity is a progressive process, $(C_1, C_2, \ldots, C_5)$follow weak order: $C_1 < C_2 < \cdots < C_5$. Set $k_0 = \min\{k | \sum_{l=k}^{5} \mu_l \ge \lambda, 1 \le k \le 5\}$, then the flight test maturity can be classified into class C_{k_0}. The confidence coefficient λ is usually between 0.6 and 0.7.

5 Case Application: Flight Test Transfer Phase Maturity Analysis on a Certain Type of Airplane

This chapter conducts research into a new type of airplane and focuses on the evaluation of its flight test maturity. Based on an extensive investigation, a comprehensive evaluation questionnaire is designed. During the evaluation process, the calculation method considering expert weight is used. A conclusion is drawn according to the analysis result of the evaluation, which indicates the rationality and validity of FTMM as well as the calculation method.

2 experts in the field of airplane design, 2 flight test experts and 1 flight test supervisor are invited for the evaluation, whose expert weight δ are 0.25, 0.25, 0.2, 0.2, 0.1 respectively.

The scores given by one of the airplane design experts on each dimension of Maturity Level 1 is shown in Table 5.

The evaluation team sets the thresholds for each dimension according to the practical situation: $k_1(1) = 0.9$, $k_2(1) = 0.9$, $k_3(1) = 0.9$, $k_4(1) = 0.9$, $k_5(1) = 0.9$.

For $j = 1, 2, 3, 4, 5$, if the inequality $\mu_j(m) \ge k_j(m)$ can be always met, the flight test maturity can be decided to have reached Level 1, and it is ready for the transfer to the next phase. While if any item in $j = 1, 2, 3, 4, 5$ cannot meet the inequality $\mu_j(m) \ge k_j(m)$, based on the cask principle, the flight test maturity can be decided

Table 5 Scores of each dimension

Dimension NO.	Total number of answers	Answers	Number of each kind of answer	Marks	Total score	Score proportion
1	9	Completely realized	6	1	8.01	89.00 %
		Partly realized	3	0.67		
		Not realized	0	0		
2	8	Completely realized	6	1	7.34	91.75 %
		Partly realized	2	0.67		
		Not realized	0	0		
3	29	Completely realized	24	1	27.35	94.31 %
		Partly realized	5	0.67		
		Not realized	0	0		
4	20	Completely realized	17	1	19.01	95.05 %
		Partly realized	3	0.67		
		Not realized	0	0		
5	27	Completely realized	22	1	25.35	93.89 %
		Partly realized	5	0.67		
		Not realized	0	0		

not to have reached Level 1, in which case some improvements should be made and another maturity evaluation should be conducted.

X represents the score proportion of each expert E_i in $(I_1, I_2, I_3, I_4, I_5)$, calculation results are as follows.

$$X = \begin{pmatrix} 89.00\,\% & 91.75\,\% & 94.31\,\% & 95.05\,\% & 93.89\,\% \\ 89.00\,\% & 95.88\,\% & 89.76\,\% & 95.05\,\% & 93.89\,\% \\ 85.22\,\% & 91.75\,\% & 93.17\,\% & 91.75\,\% & 95.11\,\% \\ 100.00\,\% & 83.38\,\% & 93.17\,\% & 95.05\,\% & 88.96\,\% \\ 88.89\,\% & 91.75\,\% & 94.31\,\% & 98.35\,\% & 100.00\,\% \end{pmatrix} \begin{matrix} E_1 \\ E_2 \\ E_3 \\ E_4 \\ E_5 \end{matrix}$$

Evaluation results are as follows.

$$\mu_1(1) = 90.43\,\%, \mu_2(1) = 90.71\,\%, \mu_3(1) = 92.91\,\%, \mu_4(1) = 95.06\,\%, \mu_5(1) = 94.40\,\%$$

It can be known that for $j = 1, 2, 3, 4, 5$, the inequality $\mu_j(m) \geq k_j(m)$ can be always met, so the flight test maturity can be decided to have reached Level 1, and it is ready for the maiden flight test.

6 Conclusion

Based on the division of test phrase in China, this chapter establishes the Flight Test Maturity Model (FTMM), which serves as the theoretical basis for the evaluation of the flight test transfer phase and changes the current situation of transfer phase evaluation featured by its strong subjectivity and the lacking in theoretical support. This method can be also applied in the transfer phase evaluation of weaponry development life cycle.

References

1. Hu YJ, Liang HT, Zhang XZ (2002) Advanced aircraft engine and the fourth generation fighter. Aero Weapon 3:45–46
2. Zhu YL (2008) Research on the maturity of space technology. Aerospace Indus Manag 5:10–13
3. CMMISM for systems engineering/software engineering, version 1.1, Staged Representation (CMMI-SE/SW, V1.1, Staged) (2002). Carnegie Mellon Software Engineering Institute
4. Wang XY (2006) Research on reliability engineering capability maturity model. Beihang University, China
5. Zhang TS (1998) Aircraft flight test manual. National Defence Industry Press, China
6. Guo DQ (2010) Maturity model and evaluation research based on the TRL technology. National University of Defense Technology, China
7. Cheng QS (1997) Attribute recognition theoretical model with application. Acta Scientiarum Naturalium Universitatis Pekinensis 33(1):12–20
8. Qiu YL (2009) The RMS management research and application of technical state management. Beihang University, China

Bayesian Optimal Design for Step-Stress Accelerated Degradation Testing Based on Gamma Process and Relative Entropy

Xiaoyang Li, Tianji Zou and Yu Fan

Abstract Accelerated degradation testing (ADT) technology for long-life and high-reliability products has become one of the key technologies in life and reliability field. The scientific and reasonable testing program can not only provide correct basis for decision making, but also make full use of resources and reduce the cost of product development. Hence, how to make full use of products' historical information to develop a short-term efficient pilot program has become a key-problem to be solved in ADT technology. This chaptere proposes the Bayesian optimal design of step-stress accelerated degradation testing (SSADT) based on Gamma process and relative entropy. Firstly, we briefly describe the applicability of Gamma process and the relative entropy in ADT, and the degradation model and relative entropy's application method are given. Secondly, under the framework of Bayesian theory, we study the Gamma degradation process based SSADT optimal design method by using maximize the relative entropy as the optimization goals and test variables as the optimization design constraints. Finally, we use a 3-steps bulb's SSADT to verify the effectiveness of the proposed method. The example shows that the method of this chapter is fast and efficient which can comprehensively use the prior information to work out the optimal pilot program.

This work was supported by the national natural science foundation of China (61104182).

X. Li (✉) · T. Zou
Science and Technology on Reliability and Environmental Engineering Laboratory, Beihang University, No. 37 Xueyuan Road, Beijing 100191, People's Republic of China
e-mail: leexy@buaa.edu.cn

Y. Fan
Beihang University, Beijing, People's Republic of China

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_82

1 Introduction

With the development of engineering and scientific technology, it is difficult to obtain enough data of failure time because many products are able to operate for a long period of time before failure. Accelerated Degradation Testing (ADT) can extrapolate product lifetime characteristics in the normal stress level by collecting product performance degradation data under high stress levels without actual failure. Therefore, ADT gets rapid development and become an international research in nearly 20 years. Weighing the mathematic theory and engineering practice two aspects, ADT optimal design gives optimal experiment scheme by scientific model of performance degradation process, reasonable description of optimization problem and efficient optimization algorithm. Hence, optimal design is the primary and most important research content of ADT technology [14].

Based on the given optimization model, traditional ADT optimal design optimizes through assuming values of parameters. But casually "assuming values of parameters" would bring the test program situations a lot of uncertainty frequently, such as over test, insufficient test, or losing the meaning of "optimal design". However, prior information, such as data from historical test or similar products before the test, is valuable objective information, which would be very useful. Thus, optimal design based on Bayesian theory gets more and more attention by academia and engineering researchers [3, 5]. Nowadays, most of researchers focus on Accelerated Life Testing (ALT), such as Erkanli and Soyer [6], Bris [1], Zhang and Meeker [28], Liu and Tang [16], Tang and Liu [24], Liu and Tang [17]. In ADT, aiming at mixed effect degradation model, Hamada et al. [10] studied the optimization design of performance degradation testing by adopting Bayesian method. By using logarithmic linear regression model to describe the performance degradation process, Liu and Tang [18] optimize the sample distribution and the stress level of CSADT, they took square loss as optimization goal of the test scheme design and fixed test cost as optimization constraints. Shi and Meeker [22] studied the Bayesian optimization design about destructive testing performance in ADT.

Gamma process is an important stochastic process. Since gamma process increment is independence and non-negativity, it is commonly used to model the degradation process with monotonous smoothly. According to the variation of parameters, gamma process can represent different degradation process and the stationary or non-stationary process. Hence, Gamma process is suitable for describing the degradation process which is random and monotonous, such as corrosion, aging, etc. Therefore, using Gamma process as ADT degradation model has an important research value [23, 25– 27].

Relative entropy is originated from the concept of information theory, which is used for the difference measurement between two different distributions. Under the framework of Bayesian theory, relative entropy can be used to measure the information gained between the prior and the posterior distribution. Compared with the asymptotic theory or D optimization method, relative entropy could comprehensively utilize sample information and prior information from the products of ADT which can

make the results much more accurately. Hence, as the goal of the experiment design optimization, relative entropy gets more and more attention [2, 4, 13, 15, 21].

Hence, this chapter considers Gamma process as the degradation model and the maximum of relative entropy as the optimization goal based on the Bayesian theory and studies testing scheme optimization design of SSADT.

2 Bayesian Accelerated Degradation Model Based on the Gamma Process

2.1 Gamma Process

Gamma process is a stochastic process whose increment is independent and non-negative with shape parameter and scale parameter [11]. As the shape parameter value changes, the shape of Gamma distribution changes and can describe different characteristic of the data and characterize the influence of stress and time on the products' performance. In addition, scale parameter describes the influence of random factors on the products' performance, such as environmental factors, human factors, subtle differences of material, etc. Hence, this article adopts Gamma process to describe the condition of degradation value varying with time.

Here, $Y(t)$ is used for representing the measured product performance value which is the difference between time t and 0; $\boldsymbol{X}$ represents the increment of degradation which is the difference value of product performance degradation between two different time.

Usually, a stochastic process $\{Y(t), t \geq 0\}$ with following characteristics is called Gamma process:

1. $Y(0) = 0$ and the probability is equal to 1
2. $Y(\tau) - Y(t) \sim \text{Gamma}(y|\alpha(\tau) - \alpha(t), \beta), \forall \tau > t \geq 0$
3. $Y(t)$ is independent increment process

$\alpha(t)$ is a non-decreasing and continuous real function, whose increment of degradation constant is greater than 0, indicating that product degradation process is irreversible. At the time 0, product performance parameters are not degraded, i.e. $\alpha(0) = 0$. Here, $\alpha(t)$ is a linear function, i.e. $\alpha(t) = ht$, and h is a constant.

Considering that product performance degradation process has a relationship with shape parameter of the Gamma process:

$$E(Y(t)) = \int_0^\infty z f_{Y(t)}(z) dz = \frac{\alpha}{\beta} = \frac{ht}{\beta} \tag{1}$$

We can see that the product performance degradation value is proportional to the t and h.

2.2 Basic Assumption

According to introduction of the Gamma process and combining with the research of SSADT base on drift Brownian motion, some basic assumptions are given below about SSADT based on Gamma process:

1. Degradation trend is monotonic, i.e., the degradation damage cannot be reversed;
2. Failure mechanisms will not change with stress levels;
3. There are no failure during ADT, i.e., product performance will not cross threshold;
4. Under the normal stress level S_0 and the accelerated stress levels, $S_1 < S_2 < \cdots < S_k$, the product performance degradation process $Y(t)$ follows the Gamma process with degradation rate $d(S_l)$ and scale parameter β, $l = 1, \ldots, k$ [11]:

$$f(x) = \frac{x^{d(S)\cdot\Delta t-1}e^{-\frac{x}{\beta}}}{\beta^{d(S)\cdot\Delta t}\Gamma(d(S)\cdot\Delta t)}. \tag{2}$$

5. The shape parameter is regarded as the function of stress conditions S_l, i.e., The shape parameter is an accelerated model and has the following log-linear expression [19, 20]:

$$\ln d(S_l) = a + b\varphi(S_l) \tag{3}$$

   where a, b are parameters needed to be estimated, $\varphi(S)$ is an known function of S.
   According to Eqs. (2) and (3), the unknown parameter vector of degradation model based on Gamma process is $\theta = (a, b, \beta)$. According to the prior information of the product, the following assumption of prior distribution is determined by the theory of Bayesian and conjugate prior.
6. Parameter vector $\boldsymbol{\theta}$ follows the following prior distributions:

$$\begin{aligned} a &\sim N(\mu_a, \sigma_a^2) \\ b &\sim N(\mu_b, \sigma_b^2) \\ \beta &\sim \exp(\mu_\beta) \end{aligned}$$

2.3 Maximum Likelihood Function of SSADT Based on Gamma Process

n samples are given under the k-level SSADT. Let S_0 be the normal stress level and, $S_1 < S_2 < \cdots < S_k$ are the accelerated stress levels. During the SSADT, there are m_l times of performance inspection on the $S_l (l = 1, \ldots, k)$ stress level; and the cumulative inspection times of the SSADT is M, i.e.

$$\sum_{l=1}^{k} m_l = M \tag{4}$$

t_{ijl} is the time of the jth inspection of the ith sample on the kth stress level ($i = 1, \ldots, n$; $l = 1, \ldots, k$; $j = 1, \ldots, m_l$). And, during the testing time $[0, t_{ilm_l}]$, inspection time in the order is:

$$t_{il1} \leq \cdots \leq t_{ilm_l}$$

After process the data, the corresponding inspection result is y_{ilj}. The increment of degradation is $x_{ilj} = y_{il\ (j+1)} - y_{ilj}$, and the inspection time interval is $\Delta t_{ilj} = t_{il\ (j+1)} - t_{ilj}$. As a result, maximum likelihood function of SSADT based on the Gamma process is:

$$p(x|\boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{l=1}^{k} \prod_{j=1}^{m_l} \left\{ \frac{x_{ilj}^{\exp[a+b\varphi(S_l)]\Delta t_{ilj}} \cdot \exp\left(-\frac{x_{ilj}}{\beta}\right)}{\beta^{\exp[a+b\varphi(S_l)]\Delta t_{ilj}} \Gamma(\exp[a+b\varphi(S_l)]\Delta t_{ilj})} \right\} \tag{5}$$

And the logarithmic likelihood function is:

$$\begin{aligned} \log p(x|\boldsymbol{\theta}) = & \sum_{i=1}^{n} \sum_{l=1}^{k} \sum_{j=1}^{m_l} \left\{ (d(S_i)\Delta t) \ln x_{ilj} - \frac{x_{ilj}}{\beta} \right\} \\ & - \sum_{i=1}^{n} \sum_{l=1}^{k} \sum_{j=1}^{m_l} \left\{ \ln \left[\beta^{d(S_i)\Delta t} \Gamma(d(S_i)\Delta t_{ilj}) \right] \right\} \end{aligned} \tag{6}$$

3 ADT Optimization Design Based on Relative Entropy

3.1 The Relative Entropy in the Bayesian Framework

Under the framework of Bayesian theory, relative entropy can be used to measure the information gained between the prior and the posterior distribution [12]. Using the Bayesian theory, the posterior distribution of a probability distribution is got, which has the new entropy different from the entropy of prior distribution. Hence, we calculate the difference of entropy to depict two probability distribution differences in the amount of information which is also called useful information.

The goal of the Bayesian optimal design is always maximization expectation of relative entropy between the prior and the posterior distribution. From the perspective of Shannon information, it represents the information gain obtained by

experiment. According to the expected information gain as the utility function obtained from the experiment come forward by Lindley, we choose the maximization of expect relative entropy as Bayesian optimization guidelines.

3.2 Optimization Objective

Based on the research in literature 21, the information I_0 contained in the prior distribution is:

$$I_0 = \int p(\boldsymbol{\theta}) \log p(\boldsymbol{\theta}) d\boldsymbol{\theta} = E_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}) \tag{7}$$

where, $p(\boldsymbol{\theta})$ is the probability density function of the prior distribution for model parameters; $E_{\boldsymbol{\theta}}$ is the expectation of $\boldsymbol{\theta}$.

The total amount of information $I_1(x)$ obtained from the posterior distribution is:

$$I_1(x) = \int p(\boldsymbol{\theta}|x, \eta) \log p(\boldsymbol{\theta}|x, \eta) d\boldsymbol{\theta} \tag{8}$$

where, $p(\boldsymbol{\theta}|\boldsymbol{x})$ is the probability density function of the posterior distribution for model parameters.

The amount of information obtained from testing program η is:

$$I(\eta, x, p(\boldsymbol{\theta})) = I_1(x) - I_0 \tag{9}$$

Since test scheme design should be given before gaining the data $\mathbf{X}$, so we need mathematical expectation of sample space information. The expectation of $I(\eta,x,p(\boldsymbol{\theta}))$ is:

$$I(\eta, p(\boldsymbol{\theta})) = E_x[I_1 - I_0] \tag{10}$$

where, E_x is the mathematical expectation of the sample space information.

Hence, the optimization goal based on relative entropy is:

$$\max_{\eta \in \mathbf{D}} I(\eta, p(\boldsymbol{\theta})) = \max_{\eta \in \mathbf{D}} E_{x,\boldsymbol{\theta}|\eta}[I_1 - I_0] \tag{11}$$

where, $I_1(x)$ is the total amount of information obtained from the posterior distribution, I_0 is the information contained in the prior distribution, E_x is the mathematical expectation of the sample space information.

Based on the research in literature [7, 9, 8], Laplace-Metropolis algorithm can be used to calculate relative entropy values effectively by Monte Carlo simulation method.

3.3 Constraints

According to the test variable scope, the constraints of the optimization problem is identified.

Constraint conditions of test variables can be divided into six parts: the total test time t; total sample size n; the number of stress levels k; the stress level, i.e. $S_0 < S_1 < S_2 < \cdots < S_k \leq S_{\max}$; allocation ratio of the test duration under each stress level, i.e. $1 > r_1 \geq r_2 \geq \cdots \geq r_k > 0$ and $\sum_{l=1}^{k} r_l = 1$, $t_l = t \times r_l$; performance monitor interval Δt.

3.3.1 Optimization Model

Combined with the constraints, all the samples are experienced each stress level by turn, i.e. $n = n_1 = n_2 = \cdots = n_k$. The optimization model is:

$$\begin{aligned} &\max E_{x,\boldsymbol{\theta}|\eta}[I_1 - I_0] \\ &s.t.\, n \geq 3 \\ &S_0 < S_1 < S_2 < \cdots < S_k \leq S_{\max} \\ &1 > r_1 \geq r_2 \geq \cdots \geq r_k > 0, \sum_{l=1}^{k} r_l = 1 \end{aligned} \tag{12}$$

Here, the variables needed to be optimized are S_l and r_l.

3.4 Optimization Algorithm

There are both continuous variables (such as the total test time t) and discrete variables (such as total sample size n) among the variables needed to be optimized in (12). These variables, the optimization objectives and the constraints constitute a complex multidimensional optimization model. Hence, optimization algorithm is an important research content. According to the engineering practice, the continuous variables can be discretized reasonably. Combining with the constraints, the set of testing scheme is determined. Using the optimization method, the target value of each testing scheme can be calculated. Then, we get the optimal solution through the enumeration method traversal optimization.

The algorithm of Bayesian optimal design for SSADT based on gamma process and relative entropy is shown in Fig. 1.

The algorithm of "calculate the relative entropy of each testing scheme in the solution set $\boldsymbol{D}$, the solution correspond to the maximum relative entropy is the optimal solution" is shown in Fig. 2.

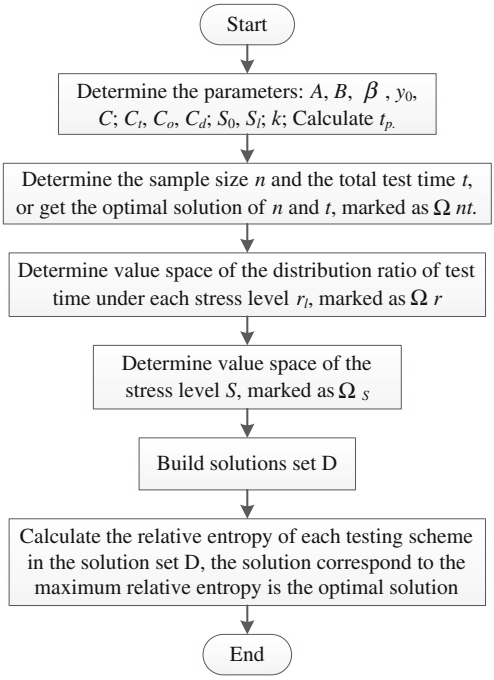


Fig. 1 Flow chart of SSADT optimization algorithm

4 Example

The resistance of bulbs would increase as the work time pass and this feature would accelerate as the voltage of bulbs rise. Hence, this section uses the ADT of bulbs to verify the proposed optimization design method.

According to the Assumption 6 and the history test information of the same bulbs, we assume the prior information of bulbs is:

$$\begin{aligned} a &\sim N(-25, 10) \\ b &\sim N(7, 10) \\ \beta &\sim \exp(10.01) \end{aligned}$$

Based on the engineering experience, the experimental parameters are shown in Table 1.

Based on the test optimization design method described in Sect. 3, we can use the test parameters and the prior information of bulbs to design the optimal test plan before the test, getting the structure of the stress value space and the inspection number value space. The stress level satisfies the uniform-interval of the stress level reciprocal, i.e. $1/S_1 - 1/S_2 = 1/S_2 - 1/S_3$, to decrease the calculation of the simulation. The inspection number satisfies $m_1 \geq m_2 \geq m_3$. When setting up the inspection number value space, we firstly determine m_1 and then $m_2 = 2/3(M - m_1)$,

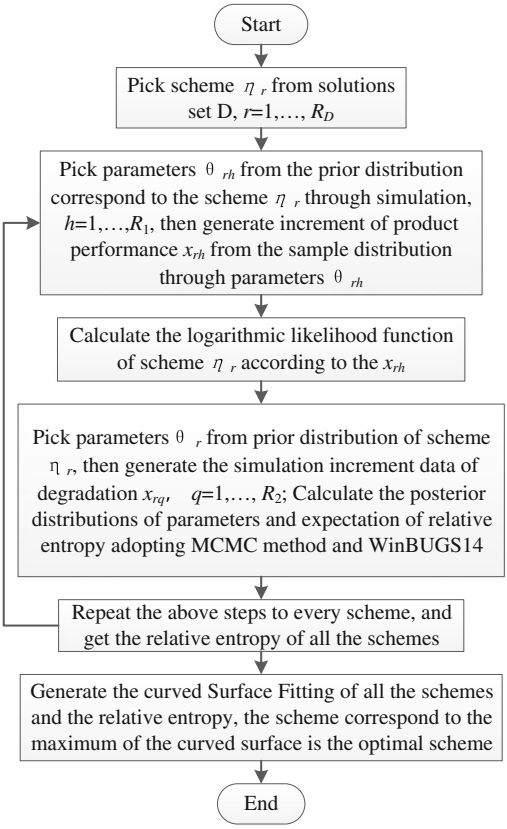


Fig. 2 Flow chart of optimization algorithm of maximum the relative entropy

Table 1.1 Test parameters

Sample size	Total number of detection	Inspection intervals (min)	Number of stress level	Low/High stress level (V)
8	80	2	3	6.3/7.5

$m_3 = M - m_1 - m_2$. Calculating the relative entropy expectations of all the schemes and using curved surface fitting to process the results, the result is shown in Fig. 3.

From Fig. 3, we know that the optimal inspection number scheme: $[m_1{:}m_2{:}m_3] = [34{:}30{:}16]$, the optimal stress level: $[S_1{:}S_2{:}S_3] = [6.3{:}6.84{:}7.5]$, and the relative entropy is 2.12×10^4.

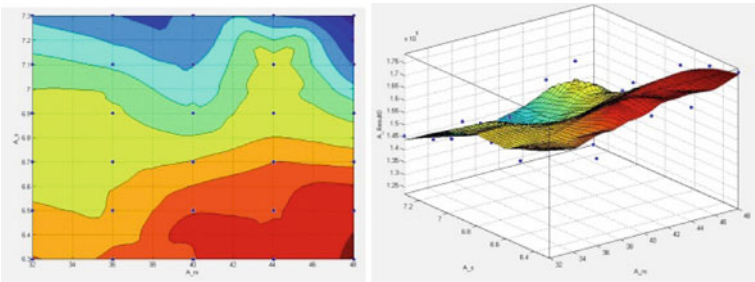


Fig. 3 Curved surface fitting of testing results

5 Conclusion

This chapter proposed the Bayesian optimal design for SSADT based on Gamma process and relative entropy. Gamma process is used as degradation model which can describe the monotonic degradation process of the product well. Meanwhile, the relative entropy is used as the optimization goal which can comprehensively represent the prior information and sample information of the product before and after the experiment. It also provides optimization design method and gets the solutions. At last, 3-step bulb's SSADT example verifies the effectiveness of the proposed method and the certain value in engineering applications.

References

1. Bris R (2000) Bayes approach in RDT using accelerated and long-term life data. Reliab Eng Syst Safety 67:9–16
2. Busetto AG, Ong CS, Buhmann JM (2009) Optimized expected information gain for nonlinear dynamical systems. In: Proceedings of the 26th annual international conference on machine learning, ACM
3. Chaloner K, Verdinelli I (1995) Bayesian experimental design: a review. Stat Sci 10 (3):273–304
4. Clarke BS, Barron AR (1990) Information-theoretic asymptotics of Bayes methods IEEE Trans Inf Theory 36(3):453–471
5. Clyde MA (2001). Experimental design: a bayesian perspective. Soc Behav Sci 1–22
6. Erkanli A, Soyer R (2000) Simulation-based designs for accelerated life tests. J Stat Plann Infer 90:335–348
7. Free Software WinBUGS. http://www.mrc-bsu.cam.ac.uk/bugs/
8. Ge Z, Li X, Jiang T (2011) Planning of CSADT with stress optimization under cost constrained. J Beijing Univ AeronautAstronaut 37(10):1277–1281 in Chinese
9. Ge Z (2012) Study on optimal design of accelerated testing based on performance degradation information. Beihang University, Beijing (In Chinese)
10. Hamada MS, Wilson AG, Shane Reese C, Martz HF (2008) Bayesian reliability. Springer Science + Business Media, New York, pp 330–331
11. http://En.wikipedia.org/wiki/Gamma_process

12. http://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence
13. Lewis SM, Raftery AE (1997) Estimating Bayes factors via posterior simulation with the laplace—metropolis estimator. J Am Stat Assoc 92(438):648–655
14. Li XY, Jiang T (2009) Optimal design for step-stress accelerated degradation testing with competing failure modes. In: Reliability and maintainability symposium, IEEE, 2009
15. Lindley DV (1956) On a measure of the information provided by an experiment. Annals Math Stat 27:986–1005
16. Liu X, Tang LC (2009) A sequential constant-stress accelerated life testing scheme and its Bayesian inference. Qual Reliab Eng Int 25(1):91–109
17. Liu X, Tang LC (2010) Planning sequential constant-stress accelerated life tests with stepwise loaded auxiliary acceleration factor. J Stat Plann Infer 140(7):1968–1985
18. Liu Xiao, Tang Loon-Ching (2010) A Bayesian optimal design for accelerated degradation tests. Qual Reliab Eng Int 26(8):863–875
19. Mao S, Wang L (2000) Accelerated life testing. Science Press, Beijing In Chinese
20. Nelson W (1990) Accelerated testing: statistical models, test plans and data analyses. Wiley Online Library
21. Qian M, Gong G, Clark JW (1991) Relative entropy and learning rules. Phys Rev A 43 (2):1061
22. Shi Y, Meeker WQ (2010) Bayesian methods for accelerated destructive degradation test planning. http://www.stat.iastate.edu/preprint/articles/2010-11.pdf
23. Si XS, Wang W, Hu CH et al (2011) Remaining useful life estimation: a review on the statistical data driven approaches. Eur J Oper Res 213(1):1–14
24. Tang LC, Liu X (2010) Planning and inference for a sequential accelerated life test. J qual Technol 42(1):103–118
25. Tsai CC, Tseng ST, Balakrishnan N (2011) Optimal burn-in policy for highly reliable products using Gamma degradation process. IEEE Trans Reliab 60(1):234–245
26. Tseng ST, Balakrishnan N, Tsai CC (2009) Optimal step-stress accelerated degradation test plan for gamma degradation processes. IEEE Trans Reliab 58(4):611–618
27. Van Noortwijk JM (2009) A survey of the application of gamma processes in maintenance. Reliab Eng Syst Safety 94(1):2–21
28. Zhang Y, Meeker WQ (2006) Bayesian methods for planning accelerated life tests. Technometrics 48(1):49–60

A Distributed Intelligent Maintenance Approach Based on Artificial Immune Systems

Marcos Zuccolotto, Luca Fasanotti, Sergio Cavalieri and Carlos Eduardo Pereira

Abstract Maintenance services logistics for wide geographically dispersed applications, such as oil transfer systems via pipelines or waste water treatment, have high costs and standard approaches usually lead to sub-optimal solutions. These systems are composed by a huge number of devices, often placed in inaccessible areas with a large distance between them. In such applications autonomous Intelligent Maintenance System (IMS) are capable to estimate their health conditions, can be used to forecast maintenance needs and to optimize maintenance schedule, therefore reducing the overall costs. Artificial Immune Systems (AIS) are a set of algorithms inspired by bio-immune systems that have features suitable for applications in IMS. AIS have distributed and parallel processing that could be useful to model large production systems. This chapter proposes an architecture for a Distributed IMS using Artificial Immune Systems concepts to face the challenges described and explore in-site learning. Each equipment has its own embedded AIS, performing a local diagnosis. If a new fault mode is detected, this information is evaluated and classified as a new non-self pattern, and included in the "vaccine". In this way, what is learned by one AIS can be propagated to the others. This proposal is modeled and implemented using multi-agent systems, where every autonomous IMS is mapped to a set of local agents, while the communication and decision process between IMSs are mapped to global agents. The chapter also describes the preliminary results deriving from the application of the proposed approach to a case study.

M. Zuccolotto (✉) · C.E. Pereira
Federal University of Rio Grande do Sul—UFRGS, Porto Alegre, Brazil
e-mail: marcos.zuccolotto@ufrgs.br

C.E. Pereira
e-mail: cpereira@ece.ufrgs.br

L. Fasanotti · S. Cavalieri
Università degli studi di Bergamo—UniBg, Bergamo, Italy
e-mail: luca.fasanotti@unibg.it

S. Cavalieri
e-mail: sergio.cavalieri@unibg.it

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_83

1 Introduction

Logistics costs related to maintenance services for wide geographically dispersed networks, such as oil transfer systems via pipelines or waste water treatment, are quite high. Standard approaches usually lead to sub-optimal solutions. Different approaches are proposed in literature to deal with these systems. Misiunas in [1] proposes a combination of reactive and proactive procedures based on Condition Based Maintenance (CBM) to manage the maintenance of water supply systems, Dey in [2] applies a risk-based Decision Support System to prioritize the right pipeline segment for inspections and maintenance. Remote monitoring could also be performed using a wireless network to enable a large area coverage [3].

These systems are composed by a huge number of devices, often placed in inaccessible areas with a large distance between them. Forecasting the maintenance needs (time to fault and required supply parts) could improve the maintenance schedule, reduce costs and ensure safe operation. In this case, autonomous Intelligent Maintenance Systems (IMS), capable to estimate the health conditions of a device without needs of human intervention, could be a useful tool to perform this forecasting [4].

IMS, such as for instance the Watchdog agent toolbox [5], include a data acquisition system and a set of algorithms, based on artificial intelligent techniques and statistical analysis, used to perform the diagnostics and/or prognostic functions. These classes of algorithms need a training stage. IMS development could use, in this stage, data acquired in field operation or reliability accelerated test. Fault modes related to other environment conditions or aging process could be hard to detect when training is performed only with laboratory data.

The Control, Automation and Robotic Group (GCAR) at Federal University of Rio Grande do Sul (UFRGS) has been developing an IMS system for control valve actuators, in a partnership with Coester, a Brazilian valve actuator manufacturer, and Transpetro, a Brazilian oil company. Research group main expertise is embedded diagnose systems, as proposed by Gonçalves et al. in [6] and Piccoli et al. in [7].

This chapter proposes an architecture for a Distributed IMS using Artificial Immune Systems (AIS) concepts. Artificial Immune Systems are a set of artificial intelligent algorithms inspired by the bio-immune system.

Immune systems are natural parallel, distributed and adaptive systems, capable to use learning, memory and associative retrieval to perform tasks as recognition and classification [8]. Analogously, an AIS aims to reproduce these features to perform tasks as pattern recognition, feature extraction, learning memory and function optimization. The use of AIS in IMS could lead to systems that are more robust, reliable and resilient [9].

Multi-Agent Systems (MAS) have been developed to cope with distribution and interoperability, performing tasks as localization of distributed information sources, integration of information in decision-making systems, in a cooperative or competitive way [10]. These features are appropriate to model the AIS proposed.

2 Artificial Immune Systems

AIS are defined by Timmis et al. as "Adaptive systems, inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving" [11].

Immune systems are a natural defense system against foreign harmful substances and microorganisms (like virus or bacteria) called pathogens. An immune system provides many levels of protection. First natural barrier against invasion is the skin, after that there is a physiological environment, where the temperature and pH establishes hostile conditions for some pathogens. Then there is the innate immune system, composed by specialized cells like macrophages, capable to identify and capture a limited set of microorganisms [12, 13].

An adaptive immune system is a more complex system, capable to identify new threats, build a response to them and embody this knowledge. AIS tries to reproduce strategies of the adaptive immune system to acquire it´s features, as distributability, adaptability, abnormality detection and disposability [12, 14].

An adaptive immune system is composed mainly of lymphocytes, the B and T cells. The recognition process is performed by chemical affinity between antibodies and molecular structures of the invaders, called antigens. Each B-cell has one particular antibody, and by a mutation process, new kind of antibodies could be generated [13, 15, 16].

The clonal selection reproduces the B-cell that was capable to identify an antigen. The B cells could be differentiated into plasma cells, that accelerate the immune response, or into memory cells, that remain more time in the organism and are responsible for the acquired immunity (learning processing) [14–16].

Lymphocytes produced by the mutation process can identify cells of the organism ("self") as invaders cells, and this full immune response can result in damage to the host organism. Negative selection is a mechanism employed to avoid this problem. It occurs within the thymus. T-cells are exposed to "self" molecular structures, and those that react against it are eliminated. The remaining T cells act as suppressor for B-cells, avoiding the recognition of a "self" structure as invaders [13–16].

The Immune Network theory proposed by Jerne and described in [17] is based on a mechanism that performs the recognition task by a network of interconnected B and T cells. These cells both stimulate and suppress each other in certain ways that lead to the stabilization of the network. So, the recognition task is performed in a system level, not as an individual task [15, 16, 18].

The developments within AIS are based on these three immunological theories, with different approaches. Clonal selection and immune networks are mainly used as learning and memory mechanisms and the negative selection principle is applied for the generation of detectors that are capable of classifying changes in self [11].

3 Using of Artificial Immune Systems in Maintenance—State of the Art

Use of AIS in maintenance have awaken an increasing interest by the research community [19, 20]. Fault detection and diagnose are performed by Clonal selection algorithms (Clonalg) in electric motors [21] and rotational machines [22, 23] and by Negative Selection Algorithms (NSA) in analog electronic systems [24] and in a DC motor [25]. Artificial Immune Network (AIN) is applied to diagnose electronic equipment [26].

Some works even propose the combination of AIS with other techniques. For instance, Jaradat and Langari in [21], proposes the use of a Fuzzy c-mean clustering to perform sensor data fusion and a AIS to detect faults in sensor systems. Thumati et al. in [27] combines an AIS with a state space observer to detect fault in nonlinear systems.

A literature survey held by the authors [20] have showed that the works that applies AIS to fault detection, diagnose or prognostic have the following aspects:

- They mainly focus on a single part or equipment, thus neglecting the natural distributed behavior of the AIS.
- AIS based fault detectors have the same or best performance than other traditional methods, but have a larger detection rate of false positives, when applied as abnormal detectors (unknown fault-modes) [23].
- Learning occurs just in the training stage or in the early operation stages [28].
- Sharing the knowledge acquired in the training stage is an approach not found in literature.

Lee et al. have introduced the idea of transforming the prognostics and health management (PHM) to engineering immune systems (EIS), in order to face the growing complexity of modern engineering systems and manufacturing process, as well as to keep these systems operating at high levels of reliability. AIS are considered the right approach to get self-maintenance systems [19].

4 Artificial Immune Intelligent Maintenance Systems

To face the PHM development challenges and to explore the research opportunities listed above, this chapter proposes an architecture for a distributed IMS that takes advantage of the distributed nature, pattern recognition and learning capabilities of the AIS, called Artificial Immune Intelligent Maintenance System (AI2MS). The AI2MS intends to provide diagnostic and prognostics of failures occurring in plant devices, instrumental in the adoption of predictive maintenance policies, with the main purpose to reducing plant downtime.

A Multi Agent System (MAS) approach has been chosen to model and implement the architecture proposed, due to the autonomous, distributed and communication features, matching the needs of the AIS and distributed applications [29, 30].

According to the MAS methodology this system is composed by several different agents, each providing a specific functionality inside the AI2MS:

- Diagnostic Agents;
- Data provider Agents;
- Prognostic Agents;
- Service Agents.

4.1 Data Provider Agents

Data provider agents regroup all the agents who are correlated to the provision of data from the field. In this category there are two different types of agents, Sensor Agents and Sensor Diagnostic Agents (Fig. 1).

Sensor Agents (*SA*): Sensor Agents are local agents which are located inside the machine and are responsible to the provisioning of field data to other agents. Each agent of this type handles a single sensor, so in a typical application there are many instances of these agents that operate at the same time, providing information to the Diagnostic Agents.

Sensor Diagnostic Agents (*SDA*): Sensor Diagnostic Agents are local agents responsible for the evaluation of the data provided by the Sensor Agents. The main task of SDA is to control the correct operation of a sensor; in case of their degradation, they can assess the fidelity of the information provided; if there is a difficulty to fix the problem generated by SA in a short time, these data could be still used.

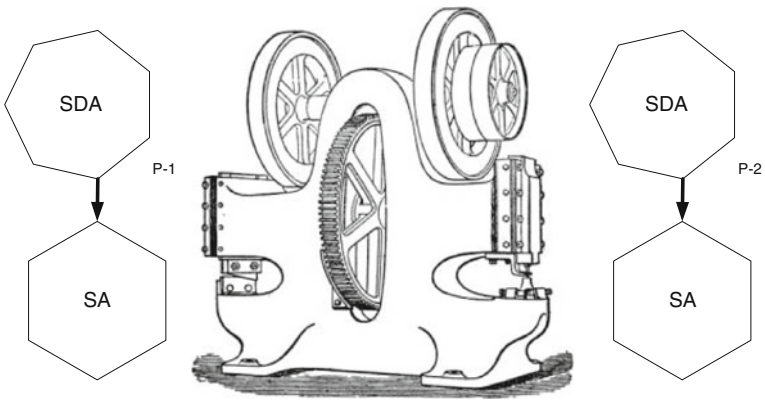


Fig. 1 Data Provider Agent

4.2 Diagnostic Agents

Within this category there are different types of agents: Fault Detection Agents (FDA), New Fault Detection Agents (NFDA) and Cooperative Detection Agents (CDA) (Fig. 2).

Fault Detection Agents (*FDA*): FDA are the kind of agents that provide fault detection capabilities of known failure modes. This type of agents represents the memory of the AIS and is the equivalent of lymphocytes T-Helper and B-Memory of a Biological Immune System.

FDA are local agents that are able to detect a specific failure mode using pattern recognition techniques; this implies that in each machine several FDA agents are continuously operative in order to increase the flexibility of the system. These agents are generated using clonal selection methodology to provide functionality of optimization of the detector and capability to detect similar failure modes.

New Fault Detection Agents (*NFDA*): The clonal selection methodology used in FDA is a good methodology to detect well known failure modes. However, whenever a new failure mode occurs, this category of agents is not effective. For this reason, in each device of the plant a specific agent is considered to detect unknown malfunction. This agent, the New Fault Detection Agent (NFDA), is based on negative selection methodology where a set of good state signatures of the device are used to train the agent making it capable to detect unknown failure modes of the system. This agent is analogous to the biological innate immune system.

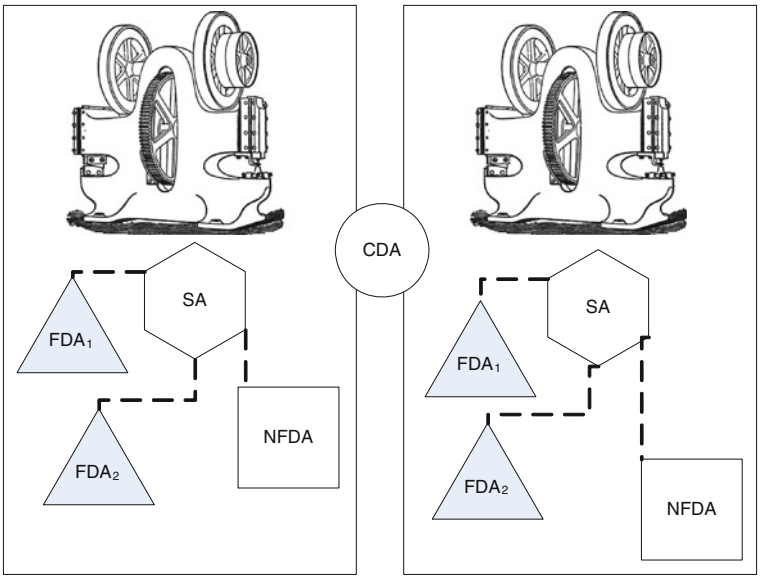


Fig. 2 Diagnostic Agents

Cooperative Detection Agents (*CDA*): Cooperative Detection Agents are network agents used to identify new fault modes (non-self), using, differently from NFDA, information provided by multiple devices. This category of agents helps the system to detect malfunctioning that are not related to a single machine but to the entire plant (e.g. a leak in a pipe). In a biological system the role of CDA is carried out by the cooperation between the different cells that constitute the immune system.

4.3 Prognostic Agents

Prognostic Agents are the group of agents responsible to provide prognostic capabilities (Fig. 3).

Device Health Assessment Agent (*DHAA*): DHAS is a local agent which runs in a single copy in each machine to estimate its remaining useful life (RUL). This agent estimates the residual health of the system using the data provided for the Sensor Agents and failure detection Agents.

Plant Health Assessment Agent (*PHAA*): Plant Health Assessment agents are similar to DHAA, acting globally through the network to estimate the health conditions of the entire plant. These agents also keep in count the topology of the plant with possible redundancies and bottlenecks.

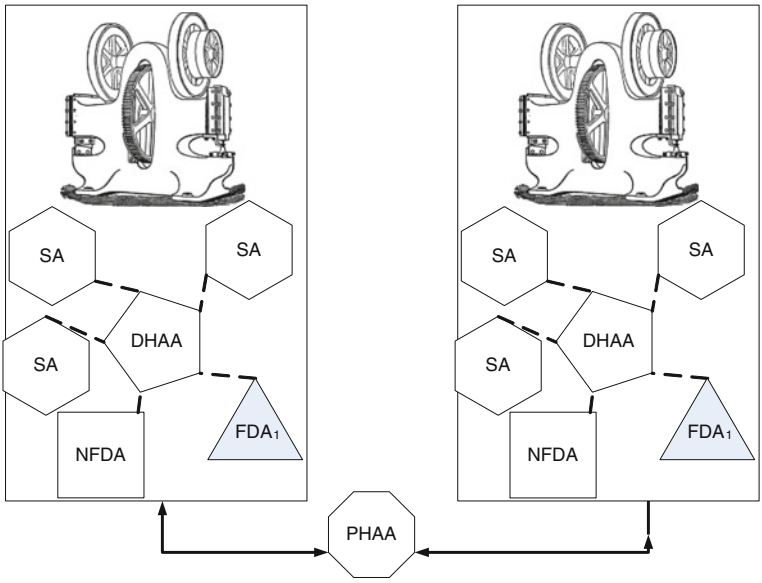


Fig. 3 Prognostic Agents

4.4 Service Agents

Service agents are a set of agents not strictly related to the diagnostic purpose. It is a set of agents responsible for the evolution and adaptation of the overall maintenance system (Figs. 4 and 5).

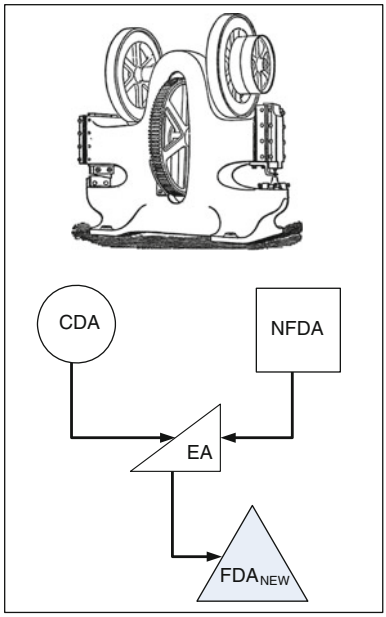


Fig. 4 Evolution Agents

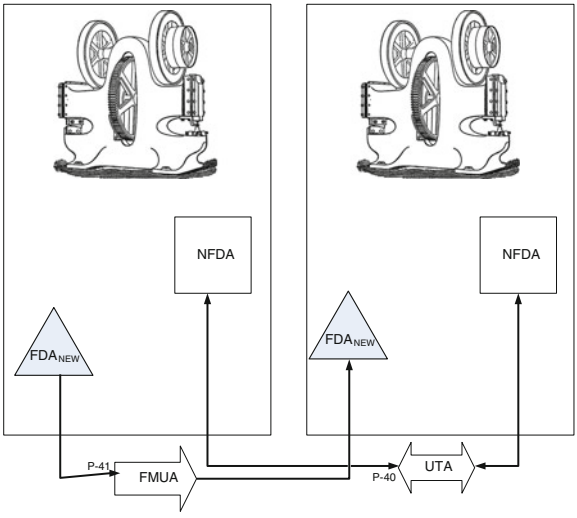


Fig. 5 Update Agents

Evolution agent (*EA*): Evolution Agents are the core of the entire AI2MS system; the role of these agents is the management of the evolution process that leads to the definition of new FDAs. This is a global agent that performs a comparison between the results of NFDA and CDA of each machine of the plants and, in cooperation with maintenance personnel, evaluate if the NFDA has really acknowledged fault mode and promote it in a FDA. This is similar to the evolution of T-lymphocytes into the thymus gland.

Failure Mode Update Agent (*FMUA*): The failure mode update Agent is a global agent responsible to share new FDA to other similar machines in order to update each machine of the plant with new detection capabilities. This update is not necessarily performed in real time, but can be made periodically in case of lack of connectivity (e.g. during a maintenance). The role of this agent is similar to the role of vaccine and colostrum in a biologic system.

Update Training Agents (*UTA*): The update training agents work in similar way respect to FMUA but share the training data for NFDA. It could mitigate the lack of training data due to few operations of one single device.

The AI2MS is planned to be scalable and adaptable to different kinds of data connection, so that system devices can operate in a standalone or connected way. Update Training agents were designed to overcome the difficulty of training due to lack of data behavior, and the association of local and global functionality should improve the fault detection and the learning capabilities of the system.

5 Case Study—Gas/Oil Pipeline

Gas/oil pipelines are plants that cover a wide geographical area, are composed by a huge number of devices, many of them of the same type, which are subject to environmental conditions, often placed in inaccessible areas. The communication between devices and the control center is usually limited. Figure 6 represents this kind of plant with two different devices, flow control valves and electric pumps.

Such a pipeline presents a large number of flow control valves (FCV); this kind of device has a reduced rate of operations. A network of FCV is an appropriate choice for application case for the AI2MS.

Each one of the FCV has its embedded AI2MS, including the Data Providers (SA + SDA), Diagnostics (FDA + NFDA) and Prognostics (DHAA). This approach

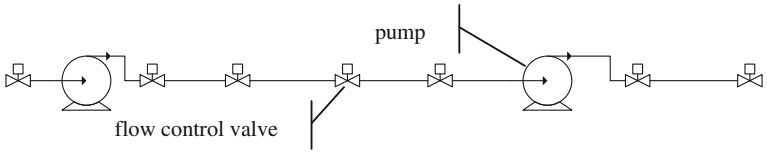


Fig. 6 A sketch-out of a Gas/Oil pipeline

is similar to other applications of AIS in IMS, as the work cited in Sect. 3 and other applications of IMS [6, 7, 31].

The global and cooperative perspective produced by the CDA, PHAA and the Services Agents leads to the AI2MS new features, like cooperative diagnostic, adaptability to environment conditions and global health assessment.

Two scenarios could be drawn in the case study, from the communications viewpoint: poor/non-existent network or good network services.

In the poor/non-existent network, functionalities that require continuous exchange of data, such as plant health assessment, cannot be provided. Migration of training data (UTA) and Fault Detection Agents can be performed when the network is online or with local connection with devices carried by personnel during preventive maintenance operation.

Operator intervention maintenance on the device can lead to the following situations:

- No specific problems.
- DHAA indicates a fault detected by FDA: correct it or schedule new intervention.
- DHAA indicates a fault detected by FDA: The intervention of the operator is needed to confirm if it is really a new fault or a different condition of a normal operation; if a new fault mode is detected, FMUA is used to propagate new acquired knowledge to other devices and EA is activated to promote the NFDA to FDA.

With the appropriate network services, cooperation between devices could be established online, besides plant health assessment and forecast of maintenance needed.

FCV in an oil/gas pipeline could operate in very different conditions from those under which the FDA has been trained, so NFDA could signal a false positive. The same NFDA signal from many devices under the same environment conditions could represent a new normal operation mode, and the CDA could help to recognize this situation and support system operation to classify this new pattern found.

Evolution of the system could be done not only finding new FDA, but testing new classification algorithms for the FDA and NFDA. Combining the efforts of UTA with the EA, it is possible to evaluate online different strategies of data processing for the Detector Agents and create a database of modes of operation and fault to improve the training of new systems.

At the present time, it was not found in literature an AIS that applies its concepts to a whole system in maintenance applications, since the majority of them are limited to one device. Possible collaboration between inter devices is generally neglected. Only recently, research works on AIS applied on smart grids, as in [32], have started exploring the challenges and opportunities deriving from such forms of interaction.

6 Conclusions and Future Work

This chapter proposes an architecture for a Distributed IMS using Artificial Immune Systems (AIS) concepts.

A short discussion about the application of AIS in maintenance activities has been carried out and some characteristics that have not yet been explored in previous works are included in this proposal.

One contribution of this chapter is the collaboration approach adopted between local AI2MS, enabling a continuous learning process. Sharing data between devices is expected to improve the performance of the whole system through the adoption of online and offline mechanisms for information exchange. The diagnostic/prognostic functionality for each single device could benefit from the past recorded behavior of other devices, subjected to same environment conditions.

Another relevant point lays in its systemic approach. AI2MS has been conceived to work in different levels, namely at: a device, inter-device and plant level. Prognostics of RUL of the plant, performed by the PHAA, can take in to account the current "health conditions" of the devices, not only relying on a static forecasting model.

AI2MS has been also planned to be scalable and highly adaptive: when the number of devices is increased, the new ones acquire, automatically, the "knowledge" of the system, by the support of Services Agents.

It is also evident that, due to the multitude of agents and the multiplicity of interaction among them, the AI2MS architecture is quite complex leading to a certain difficulty in the prediction of the behavior of the single device and, at a higher level, of the whole maintenance system. Replicating the same harmony of its natural counterpart, where it gets its inspiration, is a great challenge for future research works along this stream. In the short term, the following activities will be carried out in order to improve and foster the current version of the AI2MS system:

- Design and analysis of the intensity and quality of the message exchange, to evaluate the requirements to the network support and the possibility of integration with the plant control network.
- Incorporation of the diagnostic techniques developed by GCAR Group [6, 7] in the AI2MS.
- Evaluation of the performance of AI2MS in a simulation platform, processing real data.

Acknowledgments This work is part of a collaborative research activity between UFRGS and University of Bergamo within the ProSSaLiC Project, funded by European Community's FP7/2007–2013 under grant agreement no. PIRSES-GA-2010-269322.

Some ideas on this chapter have also been inspired by activity of UFRGS research group in collaboration with Petrobras and Coester industries.

References

1. Misiunas D (2005) Failure monitoring and asset condition assessment in water supply systems. In: Proceedings of the 7th international conference on environmental engineering 2005, (ii), pp 648–655
2. Dey PK (2004) Decision support system for inspection and maintenance: a case study of oil pipelines. IEEE Trans. Eng. Manag. 51(1):47–56
3. Zhaođxinchun Q, Wangđyapei G, Science I (2008) The application of oil and gas wells intelligent wireless monitoring system in oil field system. In: 2008 IEEE international symposium on IT in medicine and education. IEEE, pp 906–910
4. Lee J, Scott L (2006) Zero-breakdown machines and systems: productivity needs for next-generation maintenance. In: Engineering asset management 2006, pp 1–13
5. Djurdjanovic D, Lee J, Ni J (2003) Watchdog agent—an infotronics-based prognostics approach for product performance degradation assessment and prediction. Adv Eng Inf 17 (3–4):109–125 (2003)
6. Gonçalves LF, Bosa JL, Balen TR, Lubaszewki MS, Schneider EL, Henriques RVB (2011) Fault detection, diagnosis and prediction in electrical valves using self-organizing maps. J Electron Test 27:551–564
7. Piccoli LB, Henriques RVB, Schneider EL, Pereira CE (2012) Embedded systems solution for fault detection and prediction in electrical valves. In: 7th WCEAM 2012
8. Dasgupta D, Attoh-Okine N (1997) Immunity-based systems: a survey. In: 1997 IEEE international conference on systems, man, and cybernetics. Computational cybernetics and simulation, 1997 IEEE international conference on, vol 1, pp 369–374
9. Elmeligy S, Ghaffari M, Lee J (2010) Transformation from prognostics to engineering immune systems. In: Advanced maintenance engineering 2010
10. Sycara K, Decker KS, Pannu A, Willamson M, Zeng D (1996) Distributed intelligent agents. IEEE Expert 11:36–46
11. Timmis J, Hone A, Stibor T, Clark E (2008) Theoretical advances in artificial immune systems. Theor Comput Sci 403(1):11–32
12. Somayaji A, Hofmeyr S, Forrest S (1997) Principles of a computer immune system. In: Proceedings of the 1997 workshop on new security paradigms—NSPW '97. ACM Press, New York, USA, pp 75–82
13. de Castro LN, Von Zuben FJ (1999) Artificial immune systems: Part I–basic theory and applications. Universidade Estadual de Campinas
14. Dasgupta D, Forrest S (1999) Artificial immune systems in industrial applications. In: Proceedings of the second international conference on intelligent Processing and Manufacturing of Materials. IPMM'99 (Cat. No. 99EX296), vol 1. IEEE, pp 257–267
15. Aickelin U, Dasgupta D (2005) Artificial immune systems . In: Edmund K. Burke, Kendall G (eds) Search methodologies: introductory tutorials in optimization and decision support techniques. Springer, New York, pp 375–399
16. Dasgupta D (2006) Advances in artificial immune systems. IEEE Comput Intell Mag 1:40–49
17. Mizessyn F, Ishida Y (1993) Immune networks for cement plants. In: Proceedings of ISAD 93: international symposium on autonomous decentralized systems. IEEE Computer Society, pp 282–288
18. Ishiguro A, Watanabe Y, Uchikawa Y (1994) Fault diagnosis of plant systems using immune networks. In: Proceedings of the 1994 IEEE international conference on MFI '94. Multisensor fusion and integration for intelligent systems. IEEE, pp 34–42
19. Lee J, Ghaffari M, Elmeligy S (2011) Self-maintenance and engineering immune systems: towards smarter machines and manufacturing systems. Annu Rev Control 35(1):111–122
20. Zuccolotto M, Fasanotti L, Cavalieri S, Pereira CE (2013) Artificial immune systems in industrial maintenance applications: an overview. Unpublished manuscript, UniBg, UFRGS (BR)

21. Jaradat MAK, Langari R (2009) A hybrid intelligent system for fault detection and sensor fusion. Appl Soft Comput 9(1):415–422
22. Tang P, Kong H, Gan Z, Wuhan T, Chow TWS (2011) Clonal selection programming for rotational machine fault classification and diagnosis. In: Prognostics and system health management conference
23. Laurentys CA, Palhares RM, Caminhas WM (2011) A novel Artificial Immune System for fault behavior detection. Expert Syst Appl 38(6):6957–6966
24. Amaral JLM, Amaral JFM, Tanscheit R (2006) An immune fault detection system for analog circuits with automatic detector generation. In: IEEE congress on evolutionary computation. IEEE, pp 2966–2972
25. Laurentys CA, Ronacher G, Palhares RM, Caminhas WM (2010) Design of an artificial immune system for fault detection: a negative selection approach. Expert Syst Appl (Elsevier Ltd) 37(7):5507–5513
26. Hu B, Qin S (2012) Prognostic methodology for health management of electrical equipments of propulsion system in a type of vessel based on artificial immune algorithm (60875072)
27. Thumati BT, Halligan GR, Jagannathan S (2012) A novel fault diagnostics and prediction scheme using a nonlinear observer with artificial immune system as an online approximator. IEEE Trans Control Syst Technol 21:1–10
28. Gong M, Jiao L, Ma W, Ma J (2009) Intelligent multi-user detection using an artificial immune system. Sci China Ser F Inf Sci 52(12):2342–2353
29. Ramachandran B, Srivastava SK, Edrington CS, Cartes DA (2011) An intelligent auction scheme for smart grid market using a hybrid immune algorithm. IEEE Trans Ind Electron 58 (10):4603–4612
30. Dasgupta D, Yu S, Nino F (2011) Recent advances in artificial immune systems: models and applications. Appl Soft Comput 11(2):1574–87
31. Liao L, Lee J (2009) A novel method for machine performance degradation assessment based on fixed cycle features test. J Sound Vib (Elsevier) 326(3–5):894–908
32. Bhuvaneswari R, Srivastava SK, Edrington CS, Cartes DA, Subramanian S (2010) Intelligent agent based auction by economic generation scheduling for microgrid operation. In: 2010 IEEE Innovative Smart Grid Technologies Conference, pp 1–6

Towards Ontology-Based Modeling of Technical Documentation and Operation Data of the Engineering Asset

Andreas Koukias, Dražen Nadoveza and Dimitris Kiritsis

Abstract Management of engineering assets within an organization is a crucial interdisciplinary approach that aims to optimize their performance and guarantee their overall effectiveness through efficient decision making. This task is always largely supported by official technical documentation created by the asset manufacturer which describes in detail the asset's functionality, architecture as well all necessary information such as testing, operation and maintenance specifications. This valuable information has to be accessible and comprehensive since it essentially dictates the target asset configuration, operation and maintenance modes and strategies in order to guarantee the asset's performance and availability. However, current technical documentations mainly consist of textual and graphical documents that often are poorly written and constructed, misleading, unavailable, outdated and are read by users with as little effort as possible. This results in a poor connection of the operating asset with its original documentation that prevents the asset from reaching its full potential. In this work, we will propose the new concept of using ontologies as a form of documentation that accompanies the official technical documentation and is created by the manufacturer and provided to the customer. We will also propose the use of a generic asset management ontology model that asset users can be based on to create their own domain asset ontology. Finally, we will demonstrate with examples how the use of the ontology and its reasoning mechanism is ideal to identify potential problems in the operation, configuration and maintenance of the asset, as well as potentially discover areas for improvement. We expect that eventually this concept will gather all the knowledge necessary to assist in the decision making process in order to improve the asset's availability, longevity and quality of operations.

A. Koukias (✉) · D. Nadoveza · D. Kiritsis
École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
e-mail: andreas.koukias@epfl.ch

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_84

1 Introduction

Physical or engineering assets constitute the core elements of significant value to any production organization and are the backbone for its success and overall growth. Asset management in the context of engineering assets, such as machining tools and electric motors, is a challenging and crucial task. During the operation and maintenance phase of the assets' lifecycle, it aims to optimize their performance through efficient decision making and reduce their maintenance costs, increase the revenue and guarantee their overall efficiency, availability and longevity [10]. Asset management is particularly important now with the ageing of the equipment, the fluctuating requirements in the strategy and operation levels and the emphasis on health and safety requirements [2].

This task is always greatly dependent on the official technical documentation of the assets created by the asset manufacturer and provided to the user describing in detail all information concerning the assets, such as product definition and design, functionality, architecture, operating and maintenance instructions, quality assurance and safety use. Technical documentation is the compilation of these various documents which aim to describe the technical product and make available the technical know-how and product history for the subsequent users of the information such as the engineers or operators. This is why all this valuable information has to be accessible and comprehensive since it essentially dictates the target asset configuration, operation and maintenance modes as well as overall strategies in order to guarantee the assets' performance and availability within a production organisation. Especially during the warranty period of an asset, it is crucial that the technical documentation is clear and complete because otherwise there could be great losses for the asset manufacturer and in worst case even legal action against him.

However, current technical documentations mainly consist of textual and graphical documents that often are poorly written and constructed, misleading, inaccessible, incomplete, outdated and may assume knowledge that the readers don't essentially possess. Although the documentation has now greatly moved from just printed documents and booklets to electronic versions with powerful navigation and interactive content [14], thus facilitating their accessibility, many of these problems still persist. Essentially, it is still hard for the users to find specific answers to their questions and it is quite common for them to approach the documentation with as little effort as possible. This results in a poor connection of the operating asset with its original documentation that prevents the assets' overall effectiveness from reaching its full potential.

The quality of technical documentation is an important part of the perceived quality of an asset since it constitutes the first line of support for users when they encounter a problem. It is crucial that users, such as engineers, line managers and maintenance technicians, can trust it and get the information that they require, whenever needed. For example, in a simple scenario where the temperature for a specific type of equipment rises above the predefined allowed thresholds, the user may require immediately going through the equipment's exhaustive documentation

to identify the cause of the problem and check for suggested actions e.g. modify the asset's configuration or take maintenance actions. At the same time, he may also require to check for other data from other information systems within the organisation concerning the specific equipment, such as its age and maintenance history. It is clear that this can be a very time-consuming and impractical process requiring a lot of effort by the users and possible unnecessary downtime for the equipment and can thus lead to great losses for the organisation, especially in more critical situations.

This is why in this work, we will propose the new concept of using ontologies as a form of documentation that accompanies the official technical documentation and is created by the manufacturer and provided to the customer. We will propose the use of a generic asset management ontology model that manufacturers can use to create their own domain asset ontology and we will explain how the use of these ontology and the reasoning mechanism is ideal to identify potential problems in the operation, configuration and maintenance of the asset, as well as potentially discover areas for improvement.

2 Background

In this section, we will provide some background on asset management and relevant issues as well as the concept of ontologies and previous research efforts using or recommending ontologies in the asset management domain. Also, we'll provide a brief background concerning the technical documentation.

2.1 Asset Management

Asset management is a holistic and interdisciplinary approach that covers in the context of engineering assets the whole life cycle of the asset, from the acquisition to the disposal of the asset. Its scope extends from the daily operations of assets trying to meet the targeted levels of service to supporting the organization's delivery strategies, satisfying the regulatory and legal requirements and minimizing related risks and costs [2, 3, 6]. The current work focuses on the operation and maintenance phase where the aim is to optimize the overall performance of the asset and guarantee its availability and longevity.

2.2 Ontologies

Gruber [4] defines the ontologies as explicit formal specifications of the terms in a domain and relations among them. Based on the Semantic Web vision, ontologies

are proposed here since they can capture the semantics of data, resolve semantic heterogeneities, create a shared domain vocabulary and optimize data quality and availability. They possess a high degree of expressive power and formality while also providing logical reasoning mechanisms for inference of new information based on lower-order raw data and also checking for consistency, compatibility and ambiguity [13].

There are various research efforts using or recommending ontologies in the asset management domain, but to our knowledge none are focusing on the operation and maintenance phase of the asset's lifecycle. We're presenting the most representative of these efforts here. Frolov et al. [2] develop an initial and fundamental asset management ontology and subsequent process architecture in order to support an organization's asset management initiatives. Matsokis et al. [9] use ontologies with Description Logic in a case study in asset lifecycle management whereas an ontology-based implementation for exploiting the characteristics of time in asset lifecycle management systems is presented by Matsokis and Kiritsis [8]. The development of a generic asset configuration ontology is recommended by Ouertani et al. [10] in order to provide a generic and active asset configuration management framework.

2.3 Technical Documentation

Although most forms of technical documentation do not follow recognized standards, ISO has published a series of standards related to technical product documentations which are covered by ICS 01.110 [5]. After a literature review, it is obvious that currently there is a significant lack of work concerning advancements in managing and evolving technical documentation beyond the typical textual and graphical documents in written or electronic form. van Kervel [12] presents the underlying scientific theories, methodology and software technology to meet the requirements of high quality technical documentation and recommend the conceptual modeling methodology to deliver high quality enterprise models. An approach is presented by Abramovici et al. [1] where the users are provided with easy access to product information through the use of new information channels while also suggesting the integration of technical documentation into product development. In other indicative efforts are Wingkvist et al. [14, 15] show how to define and assess the quality of information of technical documentations and demonstrate with real world documentation. Tombre and Lamiroy [11] demonstrate various pattern recognition methods for querying and browsing technical documentations of all kinds. Finally, Manago et al. [7] propose a technical documentation system that is used to dynamically publish personalized content.

3 Main Proposal

3.1 Concept

In this work, we propose the new concept of using ontologies as a form of documentation that accompanies the official technical documentation and is created by the manufacturer and provided to the customer. In specific, we suggest the use of an upper asset management ontology model concerning the documentation of the operation and maintenance phases of the asset that manufacturers can use to extend and create their own domain asset ontology for their products. This ontology for a certain manufacturer can have a generic part which is common for all products as well as parts which can be further extended to be product-specific. Furthermore, we propose that the customer and eventual user of the asset also creates his own asset ontology to manage the assets of his organisation. This particular ontology would group the overall current and historical data concerning the operation, configuration, maintenance and testing of his assets while taking into consideration that the volume of this data concerning is generally high and always increasing.

The great advantage of using ontologies comes from the fact that formal rules can be added in any ontology, e.g. using a description logic language, which provide logical reasoning mechanisms and can thus be used to infer new knowledge. Taking advantage of these capabilities in our case, the manufacturer could define formal rules in his technical documentation ontology that could essentially dictate the normal or even the optimal performance of his product. More specifically, these rules could be defined for various cases such as switching to an optimal asset configuration, keeping the asset operation values within the appropriate limits, identifying abnormal behaviours and deviations, identifying the need for taking maintenance actions, locating specific hotspots for improvement of the performance or even retiring an asset etc.

After the definition of the manufacturer's technical documentation ontology along with the relevant rules and the user's asset management ontology, the next step would be to merge them in one ontology. After the merging of the two ontologies and based on the asset operation and maintenance data, these rules would be defined to infer new knowledge that would support the asset users in various cases. Of course, when the customer starts operating the asset within his own organisation, he has the capacity to add his own rules if he identifies room for further improvement or has organisation-specific requirements. The overall aim of this process is to eventually make sure that the asset behaves according to the overall specifications included in the written technical documentation. Essentially, the technical documentation ontology with rules defined by the manufacturer could be called as a form of documentation itself that could greatly facilitate the asset users in their day to day activities and as a result could gradually be used to replace the need to refer to certain aspects of the technical documentation whatsoever. The process is demonstrated in Fig. 1.

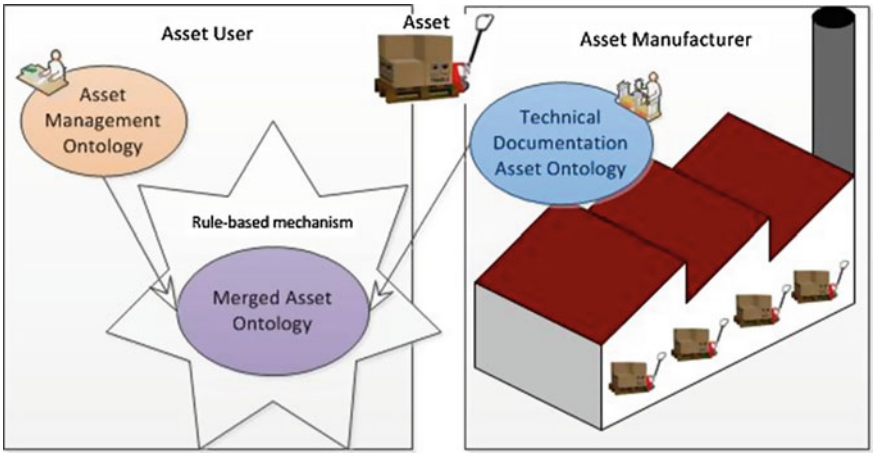


Fig. 1 Proposed concept

Furthermore, as mentioned earlier ontologies can capture the semantics of data and thus offer semantic interoperability between diverse systems. This is crucial since modern organisations use a variety of different proprietary systems to store information concerning their assets, such as SCADA and ERP. As a result, in our proposal the user of the asset does not have to perform the time-consuming process of actively searching in different data sources to link disparate data concerning the organisation assets while also consulting with the relevant technical documentation. Instead, the great benefit is that the asset ontology can integrate the data from all various sources and through the usage of the inference mechanism, it is possible to gain new knowledge that will assist in the decision making process, without essentially having to use the technical documentation itself. Besides the user, it is important to note that of course neither the manufacturer needs to be concerned about which specific asset management systems that each of his customers employs to manage his products.

In the simple scenario mentioned earlier where the temperature for a specific type of equipment rises above the predefined allowed thresholds, the proposed concept can be used to identify the best course of action. For example, a rule may be defined that will declare that if the temperature rises above 70 °C and the equipment is more than 3 years old, a maintenance action replacing a defective asset part will be recommended. A different example would be a rule that declared that in case a certain tank in the organisation has fluid level that drops below 50 cm, a maintenance action will be suggested to fix a possible leak. It is important to note in these examples that the data may originate from different systems such as SCADA (temperature, fluid level) and ERP (equipment age) and most importantly, the rules can encapsulate the content of the original technical documentation.

In the next sections, we initially present what data the user's asset management ontology should describe concerning the asset's operation and maintenance phases.

Then, we describe in a similar way the manufacturer's technical documentation ontology. Finally we briefly present the process and the benefits of merging these two ontologies.

3.2 Asset Management Ontology

In this section, we will propose the use of an upper asset management ontology model for the operation and maintenance lifecycle phases of the engineering asset, encompassing common and generic concepts which are fundamental in the asset management area. It is necessary to take into account the various domains in which the model could be applied since it should be able to adapt to various manufacturing domains and applications through extensions of the upper ontology. The so called domain ontologies created after these extensions are specific to certain manufacturing areas or activities and describe the domain in much more detail. An essential requirement to take also into consideration is that the volume of data concerning asset management e.g. for operation or configuration, is constantly increasing as time passes in dynamic manufacturing environments.

We have identified and summarized the generic information for the asset maintenance and operation lifecycle phases which correspond to various domains and should be part of any generic asset management ontology. This information can be either static e.g. asset function or dynamic e.g. asset operation data. This information is visualized in Fig. 2 and presented below:

Assets. This concerns all the engineering assets of the organisation and can be seen as the most important concept, since most of the ontology information would be connected to it. It is important to note that it may be possible to break down the asset in its technical components, which may be considered as assets themselves.
Functions. Describes the main functionality and possibly also secondary functionalities, performed by an asset e.g. refrigerating.
Processes. The process which the asset is involved in within the organisation e.g. production, filling, and is useful to identify interacting assets.
Location. This class shows the location of the asset within the organisation.
Users. The individuals in the organisation who are responsible for operating and managing an asset.
Roles. The specific actor roles that may interact with the asset e.g. line manager, quality control responsible.
Resources. The various resources that the asset may require in order to function.
States. Describes the various functioning states a certain asset can be in, such as normal state, degraded state, failure state and programmed stop state.
Operation Data. Groups the data stored during the operation of the asset, e.g. asset temperature. The instances over a period of time provide a historical view of the asset's operations.

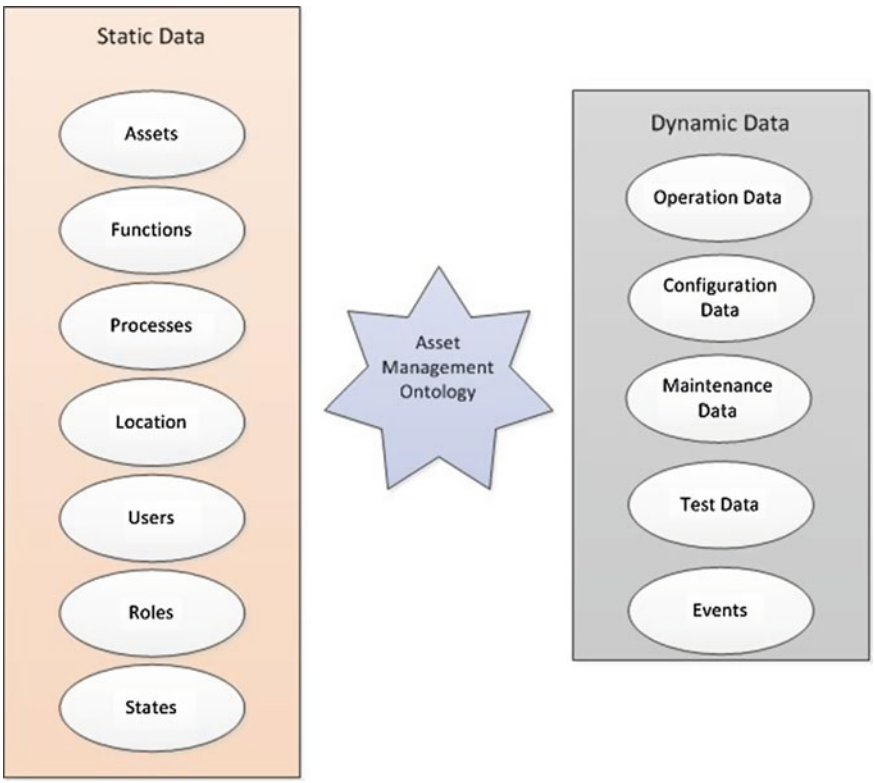


Fig. 2 Asset management ontology information

Configuration Data. Describes the records of asset configuration status at any point of time. The instances can assist in tracking the current and historical changes of asset configurations.

Maintenance Data. Contains current and historical data concerning the asset's maintenance. This concept also describes the type of maintenance type adapted by the user e.g. corrective, preventive, conditional and lists the particular sequence of actions from the available maintenance activities in a specified duration of time.

Test Data. This class groups the results of the various tests undertaken for the assets e.g. during installation or in a programmed maintenance check.

Events. This information considers the events that concern the asset. Initially, we define an event as any transient occurrence of interest for the asset which can be distinguished between internal events as changes of state caused by an internal asset transformation and external events with direct effects on the asset. In this work the low-level events are considered which are necessary for monitoring the state of the asset e.g. temperature update. The high-level events that exist on a higher abstraction level and concern the long term asset strategy

are not in the scope. The event information should also consider the alarms representing an abnormal asset's state that requires the user's attention and has warning purposes, the complex event consisting of other multiple events, the various event types, conditions and rules.

3.3 Technical Documentation Asset Ontology

The official technical documentation of the assets created by the asset manufacturer and provided to the user should describe in detail all information concerning the asset as it was originally described during the design and build phases. Although technical documentation overall considers various aspects of the asset such as product definition and design, quality assurance, safety and risk regulations are end-of-life, we are mainly interested here on the operation and maintenance aspects of the asset. The technical documentation ontology has to accurately cover all the relevant aspects of the asset while always taking into consideration the requirements of the user who may be for example a line manager or a maintenance engineer. The technical documentation ontology with its rules and capacity to infer new knowledge will essentially dictate the target asset configuration, operation and maintenance modes as well as overall strategies in order to guarantee the assets' performance and availability within a production organisation, according to the relevant technical documentation specifications. As mentioned in the asset management ontology, it is also necessary here that the ontology uses common and generic concepts that are common to various domains.

We summarize below and visualize in Fig. 3 the key generic information identified which should be part of any technical documentation asset ontology.

Operation Types. Contains the various types of operation data, as defined during the design and build phases of the asset. The various types are distinguished between functional types defining the various normal states an asset can function in and dysfunctional states where the asset's status may be degrading or may have already failed.

Configuration Types. This information groups the various types of configuration data proposed by the manufacturer. Different configuration types should correspond to different operation types and depend on certain conditions.

Test Types. Groups the various test types available for the asset which may take place e.g. during the installation of the asset or during a maintenance action.

Maintenance Strategies. This crucial information for the asset state groups the documentation maintenance information of the asset. It overall describes the various maintenance strategies suggested by the manufacturer concerning the asset and contains the various maintenance activities that the maintenance engineer can perform as part of the maintenance process of the asset e.g. replace a defective component.

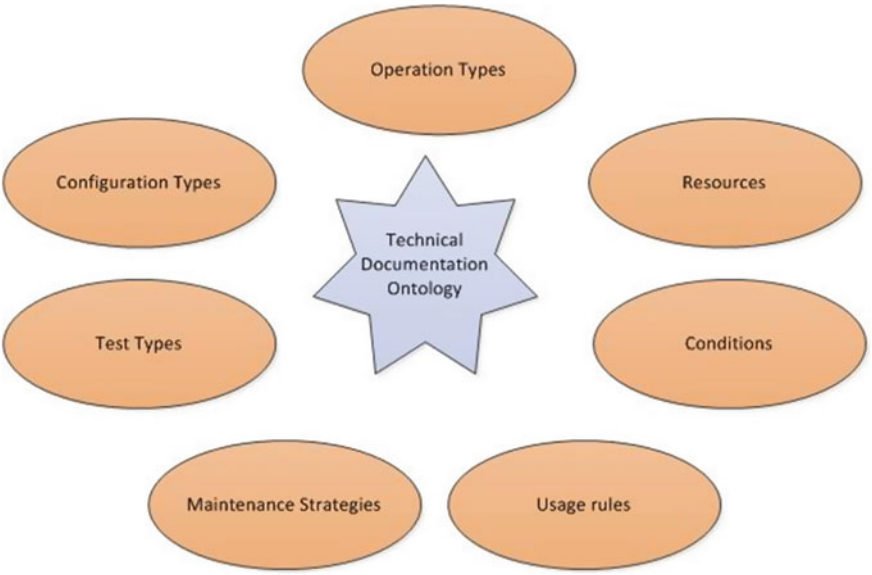


Fig. 3 Technical documentation ontology information

Usage rules. Defines some important rules in using the asset for example concerning health and safety regulations.
Conditions. Describes the conditions that are used to specify different configuration and operation modes and well as to initiate maintenance actions.
Resources. Specifies the resources that the asset requires to function in a normal operating state.

3.4 Ontology Merging

The merging of two related ontologies is obtained by taking the union of the terms and axioms defining them creating a complete new ontology. Ontology translation is the key process for the merging and is defined as the translating of datasets, queries or theories expressed using one ontology into the vocabulary supported by another. One of the core challenges of current ontology-based research is to develop efficient ontology merging algorithms which can resolve any possible mismatches with no or minimum human intervention to generate automatic a global merged ontology.

In our case, the merging of the user's asset management ontology and the manufacturer's technical documentation ontology paves the way for firstly integrating the information described previously and then for using this information and the inference mechanism to gain new knowledge. The rules initially defined in the

technical documentation ontology combined with the actual asset operating and maintenance data could now be used to find when there are divergences from the documentation specifications and thus indicate ways for identifying problems, ensuring or improving the overall performance of the asset, identifying gaps in the maintenance strategy and operation routine or even areas of improvement. The great benefit is that the information in the technical documentation ontology along with the defined rules can guarantee that the asset behaves according to the documentation specifications and as a result they could replace in large part certain aspects of the written technical and graphical documentation, thus making it unnecessary and obsolete in the decision maker's day to day activities.

4 Conclusion

This work proposed the novel concept of using ontologies as a form of technical documentation that accompanies the official technical documentation and concerns the operation and maintenance lifecycle phases of an engineering asset. When merged with a user's asset management ontology, it provides the unique capacity to use the ontologies' reasoning mechanism to identify potential problems in the operation, configuration and maintenance of the asset, as well as potentially discover areas for improvement. We expect that the proposed concept can greatly facilitate and speed up the user's decision making process while optimizing the asset's overall performance and guaranteeing its overall availability. Most importantly, we believe it can ensure that the asset behaves in a normal state according to the specifications defined in the written documentation. In the next steps, we intend to implement the ontology model in order to validate its consistency and demonstrate its benefits. We will also validate the model in a case study in order to evaluate its applicability and effectiveness on an asset's operation and maintenance phases.

References

1. Abramovici M, Krebs A, Schindler T (2013) Design for usability by ubiquitous product documentation. In: Abramovici M, Stark R (eds) Smart product engineering. Springer, Heidelberg, pp 633–641
2. Frolov V, Megel D, Bandara W, Sun Y, Ma L (2009) Building an ontology and process architecture for engineering asset management. In: 4th world congress on engineering asset, Athens, Greece
3. Frolov V, Ma L, Sun Y, Bandara W (2010) Identifying core function of asset management. In: Amadi-Echendu J et al (eds) Engineering asset management review, vol 1. Springer, London
4. Gruber TR (1993) Towards principles for the design of ontologies used for knowledge sharing. Int J Human Comput Stud 43:907–928
5. International Organisation for Standardisation. ICS 01.110: technical product documentation. http://www.iso.org/iso/catalogue_ics_browse?ICS1=01&ICS2=110&

6. Koronios A, Steenstrup C, Haider A (2009) Information and operational technologies nexus for asset lifecycle management. In: 4th world congress on engineering asset management, Athens, Greece
7. Manago M, Traphöner R, Defude B (2010) Know.Right.Now: a technical documentation system for dynamically publishing personalized content. In: Adaptivity, personalization and fusion of heterogeneous information (RIAO '10). Le Centre de Hautes Etudes Internationales D'Informatique Documentaire, Paris, France, pp 220–221
8. Matsokis A, Kiritsis D (2010) Ontology-based implementation of an advanced method for time treatment in asset lifecycle management. In: 5th world congress in engineering asset management, WCEAM 2010, Brisbane, Australia
9. Matsokis A, Zamofing S, Kiritsis D (2010) Ontology-based modeling for complex industrial asset lifecycle management: a case study. In: 7th international conference on product lifecycle management, PLM'10, Bremen, Germany
10. Ouertani MZ, Srinivasan V, Parlikad AKN, Luyer E, McFarlane DC (2009) Through-life active asset configuration management. In: International conference on product lifecycle management, PLM' 09, PLM-SP5—2009 Proceedings, Bath, UK, pp 119–207
11. Tombre K, Lamiroy B (2008) Pattern recognition methods for querying and browsing technical documentation. In: Ruiz-Shulcloper J, Kropatsch W (eds) Progress in pattern recognition, image analysis and applications. LNCS, vol 5197. Springer, Heidelberg, pp 504–518
12. van Kervel SJH (2011) High quality technical documentation for large industrial plants using an enterprise engineering and conceptual modeling based software solution. In: De Troyer O, Bauzer Medeiros C, Billen R, Hallot P, Simitsis A, Van Mingroot H (eds) ER workshops 2011. LNCS, vol 6999. Springer, Heidelberg, pp 383–388
13. Wang XH, Zhang DQ, Gu T, Pung HK (2004) Ontology based context modeling and reasoning using OWL. In: Pervasive computing and communications workshops, 2004. Proceedings of the second IEEE annual conference on. IEEE, pp 18–22, Mar 2004
14. Wingkvist A, Ericsson M, Lincke R, Löwe W (2010) A metrics-based approach to technical documentation quality. In: Proceedings of the 7th international conference on the quality of information and communications technology (QUATIC'2010)
15. Wingkvist A, Löwe W, Ericsson M et al (2010) Analysis and visualization of information quality of technical documentation. In: 4th European conference on information management and evaluation, 2010, pp 388–396

Optimal Policy Study on Reliability-Centered Preventive Maintenance for a Single-Equipment System

Q.M. Liu, M. Dong and W.Y. Lv

Abstract Because of demanding higher operational efficiency and safety in industrial systems, system maintenance trends to the direction of high speed, high load and high automation. Reliability-centered preventive maintenance schedule plays an important role in production, but it is always a complex task to make such a plan. In this paper, the optimization of reliability-centered preventive maintenance (PM) for a single-equipment system can be proposed. In order to reduce downtime loss, several PM actions are performed together using the threshold for opportunity PM. By using Markov decision process method, the transition probability matrix for different PM actions can be obtained, and then the optimal combination problem on PM strategy can be developed. Finally, the optimal PM schedule which can provide the desired levels of reliability to a single-component system and minimize maintenance cost can be obtained. A case is used to demonstrate the implementation and potential applications of the proposed method.

1 Introduction

To keep a system in normal condition, taking proper maintenance becomes even more important during its life. Many effective equipment maintenance strategies have been developed [1–4]. Generally, the equipment maintenance can be classified into corrective maintenance (CM) and PM. Normally, PM is more effective than CM because it is always to keep a system in an available condition so that the large loss caused by unpredictable fails can be avoided.

Q.M. Liu · M. Dong (✉)
Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200240, China
e-mail: mdong@sjtu.edu.cn

W.Y. Lv
Business School, University of Shanghai for Science and Technology, Shanghai 200240, China

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_85

The CM involves the repair or replacement of components which have failed or broken down. When corrective maintenance is carried out after a failure, usually the equipment can be restored to an operational condition in which it can perform its intended functions. Kenne and Nkeungoue [5] introduced the CM and PM into the production systems, proposed a joint model of production, corrective maintenance and preventive maintenance. The preventive maintenance is a schedule of planned maintenance actions aimed at the prevention of equipment breakdowns and failures. It is designed to preserve and enhance equipment reliability by replacing worn components before they actually fail. Fitouhi and Nourelfath [6] dealt with the problem of integrating non-cyclical PM and tactical production planning for a single machine. Bartholomew-Biggs et al. [7] considered the optimal PM scheduling and dealt with the problem of scheduling imperfect PM for equipment. Cassady and Kutanoglu [8] proposed an integrated model that coordinates PM planning decisions with single-machine scheduling decisions so that the total expected weighted completion time of jobs was minimized. Most of them always concentrated on the development of mathematical models in achieving the optimization of PM policy based on some specific supporting, such as uniform improvement, maintenance activity and cost, etc. For a system which is consisted of many components, the effectiveness of maintenance mainly depends on both the improved levels and the maintenance costs of the components.

For scheduling the PM, based on the defined actions, the maintenance time and the cost of a system would affect the contents of actions adopted. Considering the time of PM taken, PM policies can be classified into two kinds, periodical PM and non-periodical PM [9, 10]. In this paper, for a single-equipment system, the downtime loss could be obviously reduced as well as its effectiveness can be promoted if its reliability can be set or maintained at someone level. Reliability and maintenance cost are adopted as a criterion for scheduling maintenance actions. Thus, reliability under PM is given. Different kinds of action are taken on each stage. The maintenance model considering different maintenance actions and reliability is proposed, and reliability-centered PM strategy is scheduled. The purpose is not only on maintaining the system life to its expected life but also in obtaining the maximum system benefit by reliability optimization. By case study, the results show that the maintenance which considers more than one action is more advantage than that only single action adopted.

2 Preventive Maintenance Actions

In this paper, various maintenance actions can be adopted to slow down the system degradation. The system life will increase with time and will be affected by the maintenance actions. For time interval $[0, T]$ ($[0, T]$ can be divided into n time interval, $n \in N$, $\Delta T = T/n = T_p = (t_{p-1}, t_p)$.), the available maintenance actions set can be expressed as follows:

$$PM_i = \{PM_0, PM_1, PM_2, PM_3\}$$

where, PM_0: No maintenance action is adopted. PM_1: Adopt some daily non-replacement maintenance actions (mechanical service). PM_2: Adopt some minor non-replacement maintenance actions (repair). PM_3: Adopt replacement maintenance actions by directly using new parts to replace the old ones.

In order to develop the reliability-centered preventive maintenance optimized strategy for a single-equipment system, some basic descriptions for the PM are described as follows.

(1) M components of equipment S will all be the new status at time 0. The failure distributions among M components are independent, and are subject to Weibull distribution.
(2) When the component reaches the minimum reliability, PM_2 or PM_3 will be adopted. Otherwise, minor repairs will be adopted. For time interval [0, T], PM_1 can be adopted for components that don't reach the minimum reliability. The maintenance time of minor repairs and PM_1 can be ignored.
(3) The PM triggered component j ($j \neq i$) of component i at time interval (t_p, $\varepsilon\Delta T + t_p$) ($w \in N$) adopts the opportunity preventive maintenance.

3 Reliability under Preventive Maintenance

With adopting the PM, the failure parts can be repaired, and the surviving parts can be improved. Thus, the reliability of a system can be improved, and the reliability of the component i at time t can be denoted as follows:

$$R_i(t) = R_{i.k,s} \times R_{v,k}(t)$$

where, $R_{i,k,s}$ is the initial reliability of the failure parts after the kth PM for component i, and $R_{v,k}$ is the reliability of the surviving parts after the kth PM for component i.

Based on the different maintenance action of the kth PM, the reliability can be divided into three categories after adopting the kth PM for component i. Considering preventive maintenance whose interval is [t_p, t_p'], the reliability of the failure parts and surviving parts can be obtained as follows:

(1) $PM_1 : R_{i.k,s} = R_{i,k-1,f}, \quad R_{v,k}(t) = \exp\left(-\left(\frac{t-t_p}{m_2\alpha(i)}\right)^{\beta(i)}\right)$

(2) $PM_2 : R_{i.k,s} = R_{i,k-1,f} + m_1(R_{i.k-1,s} - R_{i,k-1,f}), \quad R_{v,k}(t) = \exp\left(-\left(\frac{t-t_p}{m_2\alpha(i)}\right)^{\beta(i)}\right)$

(3) $PM_3 : R_{i.k,s} = 1, \quad R_{v,k}(t) = \exp\left(-\left(\frac{t-t_p}{\alpha(i)}\right)^{\beta(i)}\right)$

Thus, the reliability of the component i can be obtained as follows:

$$R_i(t) = R_{i,k,s} \exp\left(-\left(\frac{t - t_p}{m\alpha(i)}\right)^{\beta(i)}\right), \quad m = \begin{cases} 1 & X_i(T_p) = PM_1 \\ m_2 & X_i(T_p) = PM_2, PM_3 \end{cases}$$

where, $R_{i,k,f}$ is the finishing reliability of failure parts before the k + 1th PM for component i. m_1 and m_2 are the improvement factors ($0 < m_1, m_2 < 1$) [9]. $X_i(T_p)$ represents the maintenance action at the time interval T_p for component i. $\alpha(i)$ and $\beta(i)$ are scale parameter and shape parameter of Weibull distribution, respectively.

4 Integrated Decision Model

4.1 Maintenance Cost

The PM process of equipment S at time interval [0, T] is Markov decision process. If the k-th PM can be adopted for component i at time T_S, assuming that $X_i(T_S) = A$ ($A = PM_1, PM_2, PM_3$). Then, the transition probability of $X_i(T_{p+S}) = B$ ($B = PM_0, PM_1, PM_2, PM_3$) is $p(X_i(T_{p+S}) = B | X_i(T_S) = A)$ at time interval T_{p+S} ($p = 1, 2, \ldots n$). The transition probability depends on the average maintenance cost per unit time when the component reaches the next minimum reliability ($R_{\min}(i)$) from taking PM actions.

Let $sg_i(T_p)$ denote the adopting PM_2 or PM_3 at time interval T_p, and $V(T_p)$ be the number of component adopting PM_2 or PM_3 at T_p, and $op_i(T_p)$ describe the adopting opportunity PM_2 or PM_3 at T_p. $R_i(t)$ represents the reliability of component i at time t.

$$sg_i(T_p) = \begin{cases} 1 & R_i(T_{p+1}) < R_{\min}(t) \\ 0 & R_i(T_{p+1}) \geq R_{\min}(t) \end{cases}, \quad V(T_p) = \sum_{i=1}^{M} sg_i(T_p),$$

$$op_i(T_p) = \begin{cases} 1 & R_i(T_{p+\varepsilon'+1}) < R_{\min}(t) \\ 0 & R_i(T_{p+\varepsilon'+1}) \geq R_{\min}(t) \end{cases}$$

In this paper, let $C_{B,i}(k + 1)$ denote the average maintenance cost per unit time at time interval [t_{p+S}, $t_{R\min}$] after adopting maintenance B ($V(T_{p+S}) \geq 1$), and it can be shown as follows:

$$C_{B,i}(k+1) = \begin{cases} C_{0,i}(k+1) = \begin{cases} C_{1,i}(k) & X_i(T_S) = PM_1 \\ C_{2,i}(k) & X_i(T_S) = PM_2 \\ C_{3,i}(k) & X_i(T_S) = PM_3 \end{cases} & B = PM_0 \\ C_{1,i}(k+1) = \frac{C_b(i) + C_m(i) F_{1,i}(k+1)}{t_{R_{\min}} - t_{p+S}} & B = PM_1 \\ C_{2,i}(k+1) = \frac{C_x(i) + C_d D_{2,i}(T_{p+S}) + C_m(i) F_{2,i}(k+1)}{t_{R_{\min}} - t_{p+S}} & B = PM_2 \\ C_{3,i}(k+1) = \frac{C_g(i) + C_d D_{3,i}(T_{p+S}) + C_m(i) F_{3,i}(k+1)}{t_{R_{\min}} - t_{p+S}} & B = PM_3 \end{cases} \quad (1)$$

$$F_{\omega,i}(k+1) = \int_{t_{p+S}}^{t_{R_{\min}}} \lambda_{\omega,k+1}(i,t)dt \quad \omega = 0,1,2,3$$

$$t_{R_{\min}} = m\left(\exp\left(\frac{1}{\beta(i)}\log\left(-\log\left(\frac{R_{\min}(i)}{R_{i,k+1,s}}\right)\right)\right)\alpha(i) + \frac{1}{m}t_{p+S}\right)$$

$$D_{\varphi,i}(T_{p+s}) = \begin{cases} \rho_B(i) - \sum_{j=1}^{i-1} D_{\varphi,i}(T_{p+s}) & \\ 0 & \varphi = 2,3, B = 2,3 \end{cases}$$

where, $C_b(i)$, $C_x(i)$ and $C_g(i)$ are the cost of adopting one PM_1, PM_2, PM_3, respectively. Cm is the cost of adopting one minor repair. C_d is the downtime cost per unit time. $D_{2,i}(T_{p+S})$ and $D_{3,i}(T_{p+S})$ represent downtime for component i at T_p adopting PM_2 and PM_3, respectively. $t_{R\min}$ is the time when the component i reaches the next minimum reliability ($R_{\min}(i)$) from taking PM actions at T_p. $F_{1,i}(k+1)$, $F_{2,i}(k+1)$ and $F_{3,i}(k+1)$ represent the number of failure for component i at $[t_{p+S}, t_{R\min}]$ adopting PM_1, PM_2, PM_3, respectively. $\rho_1(i)$ and $\rho_2(i)$ are the average maintenance time for component i adopting PM_2, PM_3, respectively.

The transition probability can be obtained as follows:

(1) $sg_i(T_{p+S}) = 0$ and $op_i(T_{p+S}) = 0$ (Adopting PM_0 or PM_1 for component i)

$$p(X_i(T_{p+S}) = b | X_i(T_S) = A) = \frac{1/C_{b,i}(k+1)}{1/C_{0,i}(k+1) + 1/C_{1,i}(k+1)} \quad b = 0,1$$

(2) $sg_i(T_{p+S}) = 1$ or $op_i(T_{p+S}) = 1$ (Adopting PM_2 or PM_3 for component i)

$$p(X_i(T_{p+S}) = b | X_i(T_S) = A) = \frac{1/C_{2,i}(k+1)}{1/C_{2,i}(k+1) + 1/C_{3,i}(k+1)} \quad b = 2,3$$

4.2 Maintenance Model

Based on Eq. (1), for the PM cost, minor maintenance cost and downtime cost, the maintenance model can be obtained as follows:

$$
\begin{aligned}
\min C = & \sum_{i=1}^{M}\sum_{p=1}^{n} X_i(T_p)\left(\frac{1}{3}C_b(i)\varpi_{1,i}(T_p) + \frac{1}{2}C_x(i)\varpi_{2,i}(T_p) + C_g(i)\varpi_{3,i}(T_p)\right) \\
& + \sum_{i=1}^{M} C_m(i)\left(\sum_{k=0}^{K(i)-1}\int_{t_{x,i}(k)}^{t_{x,i}(k+1)} \lambda_k(i,t)dt + \int_{t_{x,i}(K(i))}^{T} \lambda_{K(i)}(i,t)dt\right) + \sum_{p=1}^{n} C_d\left(\sum_{i=1}^{M} D_{\varphi,i}(T_p)\right) \\
& s.t. \\
& \varpi_{a,i}(T_p) = \begin{cases} 1 & X_i(T_p) = a \\ 0 & X_i(T_p) \neq a \end{cases} \quad a = PM_1, PM_2, PM_3 \\
& \left[\sum_{i=1}^{M} X_i(T_{p2})\right]\left[\sum_{i=1}^{M} X_i(T_{p3})\right] = 0 \\
& \forall i, t, \exists R_i(t) \geq R_{\min}(i) \\
& 0 \leq \left|t_{q2} - t_{q3}\right| \leq \varepsilon \Delta T
\end{aligned}
\tag{2}
$$

where, $K(i)$ represents the total number of PM for component i at [0, T]. $t_{x,i}(k)$ describes the time adopting the kth PM for component i.

5 Case Study

To validate the proposed methods for PM, a case is studied. A single-equipment system consists of 6 components, and each component works from new status. The time interval of equipment S is [0, 365], and $\Delta T = 1d$. The downtime cost adopting PM_2 and PM_3 is 550/d. The improvement factors are 0.8 ($m_1 = m_2 = 0.8$). The opportunity PM threshold ε is $3d$. Parameters can be obtained in Table 1.

First, each component separately adopts reliability-centered preventive PM_2 or PM_3. The maintenance cost can be obtained in Table 2.

Then, based on Eq. (2), the PM strategy optimization model can be obtained for equipment within one year. The results can be shown in Table 3. It can be seen from Table 3 that the proposed method has a better performance for saving maintenance cost. And the accuracy will be better with the growth of simulation times. The result shows that the total maintenance cost by the proposed method is 32906.71.

Table 1 Parameter values of the components

i	$C_b(i)$	$C_x(i)$	$C_g(i)$	$C_m(i)$	α(i)/d	β(i)	$\rho_1(i)$	$\rho_2(i)$	$R_{min}(i)$
1	40	100	320	480	52	1.6	0.2	0.4	0.7
2	40	220	470	1000	115	2.7	0.4	0.6	0.75
3	70	320	700	1500	123	3.4	0.3	0.5	0.8
4	30	100	220	280	165	3.3	0.1	0.3	0.6
5	10	50	70	220	35	2.2	0.4	0.6	0.6
6	70	260	720	1120	85	1.7	0.5	0.9	0.8

Table 2 Maintenance cost for each component

i	1	2	3	4	5	6	Total
C	7974.88	6106.04	6028.26	1572.16	11446.72	17930.96	51059.02

Table 3 Computed results based on the proposed method

Simulation time	Minimum total cost	Decreased cost (%)
1,000	33415.14	34.56
10,000	33296.62	34.79
20,000	32906.71	35.56

Table 4 Optimal preventive maintenance schedule

Time (d)	Component i					
	1	2	3	4	5	6
15	PM_1	PM_1	PM_1	PM_0	PM_3	PM_1
45	PM_3	PM_0	PM_1	PM_1	PM_3	PM_1
90	PM_1	PM_0	PM_1	PM_0	PM_3	PM_3
120	PM_1	PM_1	PM_1	PM_0	PM_2	PM_3
145	PM_1	PM_1	PM_1	PM_0	PM_3	PM_1
198	PM_1	PM_1	PM_1	PM_0	PM_2	PM_1
212	PM_1	PM_3	PM_1	PM_0	PM_1	PM_0
233	PM_0	PM_1	PM_1	PM_0	PM_1	PM_3
260	PM_1	PM_1	PM_0	PM_0	PM_2	PM_2
269	PM_0	PM_1	PM_0	PM_1	PM_3	PM_1
297	PM_1	PM_1	PM_0	PM_1	PM_2	PM_1
359	PM_1	PM_0	PM_0	PM_0	PM_2	PM_1

Finally, based on the proposed method, the optimal preventive maintenance schedule can be obtained, and it can be shown in Table 4. It can be seen that component 4 adopts PM_0, components 1, 2, 3 adopt PM_1, component 5 adopts PM_2 and component 6 adopts PM_3 at time 120d. Component 5 adopts the maximum number of PM_2 and PM_3, however, component 4 only adopts PM_1 to reduce failures and keep the reliability of a system.

6 Conclusions

This paper presents a reliability-centered PM strategy for a single-equipment system. The possible maintenance actions are classified into three types (PM_1, PM_2 and PM_3) which are concurrently considered on every PM stage, and minor repairs are adopted for the sudden failure. In this paper, first, based on the improvements of the

surviving and failure parts for constructing the reliability model, the effects of maintenance to reliability are formulated. Then, in order to reduce the downtime cost, multiple maintenance actions are integrated based on opportunity maintenance threshold. By using Markov decision process, PM optimization problems for a single-equipment system can be effectively solved at [0, T]. Finally, the optimal maintenance strategy can be obtained, and the dynamic optimization PM maintenance schedule can be developed.

References

1. Cekyay B, Ozekici S (2012) Optimal maintenance of systems with Markovian mission and deterioration. Eur J Oper Res 219(1):123–133
2. Huynh KT, Castro IT, Barros A, Bérenguer C (2012) Modeling age-based maintenance strategies with minimal repairs for systems subject to competing failure modes due to degradation and shocks. Eur J Oper Res 218(1):140–151
3. Liao GL, Sheu SH (2011) Economic production quantity model for randomly failing production process with minimal repair and imperfect maintenance. Int J Prod Econ 130 (1):118–124
4. Yang SL, Ma Y, Xu DL, Yang JB (2011) Minimizing total completion time on a single machine with flexible maintenance activity. Comput Oper Res 38(4):755–770
5. Kenne JP, Nkeungoue LJ (2008) Simultaneous control of production, preventive and corrective maintenance rates of a failure-prone manufacturing system. Appl Numer Math 58 (2):180–194
6. Fitouhi MC, Nourelfath M (2012) Integrating noncyclical preventive maintenance scheduling and production planning for a single machine. Int J Prod Econ 136(1):344–351
7. Bartholomew-Biggs M, Zuo MJ, Li XH (2009) Modeling and optimizing sequential imperfect preventive maintenance. Reliab Engi Syst Saf 94(1):53–62
8. Cassady CR, Kutanoglu E (2005) Integrating preventive maintenance planning and production scheduling for a single machine. IEEE Trans Reliab 54(2):304–310
9. Tsai YT, Wang KS, Tsai LC (2004) A study of availability-centered preventive maintenance for multi -component systems. Reliab Eng Syst Saf 84(3):261–270
10. Martorel S, Sanchez A, Carlos S, Serradell V (2002) Comparing effective-ness and efficiency in technical specifications and maintenance optimization. Reliab Eng Syst Saf 77(3):9–281

Optimal Burn-in Policy for Highly Reliable Products Using Inverse Gaussian Degradation Process

Mimi Zhang, Zhisheng Ye and Min Xie

Abstract Burn-in test is a manufacturing procedure implemented to identify and eliminate units with infant mortality before they are shipped to the customers. The traditional burn-in test, collecting event data over a short period of time, is rather inefficient. This problem can be solved if there is a suitable quality characteristic (QC) whose degradation over time can be related to the lifetime of the product. Optimal burn-in policies have been discussed in the literature assuming that the underlying degradation path follows a Wiener process or a gamma process. However, the degradation paths of many products may be more appropriately modeled by an inverse Gaussian process which exhibits a monotone increasing pattern. Here, motivated by the numerous merits of the inverse Gaussian process, we first propose a mixed inverse Gaussian process to describe the degradation paths of the products. Next, we present a decision rule for classifying a unit as typical or weak. A cost model is used to determine the optimal burn-in duration and the optimal cut-off level. A simulation study is carried out to illustrate the proposed procedure.

Keywords Burn-in test · Mixture distribution · Inverse Gaussian process

M. Zhang (✉) · M. Xie
Department of Systems Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong
e-mail: mmzhang5-c@cityu.edu.hk

M. Xie
e-mail: minxie@cityu.edu.hk

Z. Ye
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
e-mail: iseyez@gmail.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_86

1 Introduction

An inverse Gaussian process is a stochastic process with monotone increasing paths. The inverse Gaussian process was proposed by Wasan [1] and further studied by Wang and Xu [2] and Ye and Chen [3]. Wang and Xu [2] proposed a method to incorporate random effects in the inverse Gaussian process; Ye and Chen [3] showed that the inverse Gaussian process is a limiting compound Poisson process. In the context of the inverse Gaussian process, the probability density function and cumulative distribution function of the first hitting time to a fixed threshold have closed forms. By contrast, the increments of a Gamma process do not bear a closed-form cumulative distribution function, which makes the ensuing statistical inference intractable. This chapter makes an investigation into a burn-in test in which the degradation of the items in a heterogeneous population is modelled by a mixture inverse Gaussian process. A cost model is used to determine the optimal cut-off point and the optimal termination time.

Recall that an inverse Gaussian distribution with mean $u > 0$ and shape parameter $\lambda > 0$, denoted to be $IG(u, \lambda)$, has probability density function [4]

$$f_{IG}(x; u, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{-\frac{\lambda(x-u)^2}{2u^2 x}\right\}, \quad x > 0,$$

and cumulative distribution function

$$F_{IG}(x; u, \lambda) = \Phi\left[\sqrt{\frac{\lambda}{x}}\left(\frac{x}{u} - 1\right)\right] + \exp\left(\frac{2\lambda}{u}\right)\Phi\left[-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{u} + 1\right)\right], \quad x > 0,$$

in which $\Phi(\cdot)$ is the standard normal cumulative distribution function. An inverse Gaussian process $\{X(t), t \geq 0\}$ is a continuous-time stochastic process with the following properties:

- $X(t)$ has independent increments; that is, for all $t_2 > t_1 \geq s_2 > s_1 \geq 0$, the increments $X(t_2) - X(t_1)$ and $X(s_2) - X(s_1)$ are independent.
- $\mathrm{X(t)} - \mathrm{X(s)}$ follows an inverse Gaussian distribution $IG(\Lambda(t) - \Lambda(s), \eta[\Lambda(t) - \Lambda(s)]^2)$, for all $t > s \geq 0$.

As with the convention, let $\Lambda(t)$ be a non-decreasing, right-continuous, real-valued function for $t \geq 0$, with $\Lambda(0) = 0$. Hence, X(t) is a random variable having distribution: $IG\left(\Lambda(t), \eta\Lambda(t)^2\right)$ with mean $\Lambda(t)$ and variance $\Lambda(t)/\eta$.

In the framework of condition-based maintenance, a monitored product's failure can be suitably defined as the event of which the product's deterioration level first reaches a pre-determined threshold D. Let X(t) denote the deterioration of the product at time $t \geq 0$ and T_D the first-passage time of the degradation to the threshold D. Due to the monotonicity of the degradation path of the inverse

Gaussian process, the cumulative distribution function of T_D can be readily obtained:

$$P(T_D \le t) = P(X(t) \ge D) = 1 - P(X(t) < D) = 1 - F_{IG}\left(D; \Lambda(t), \eta\Lambda(t)^2\right).$$

Specifically, we have

$$P(T_D \le t) = \Phi\left[\sqrt{\frac{\eta}{D}}(\Lambda(t) - D)\right] + \exp(2\eta\Lambda(t))\Phi\left[-\sqrt{\frac{\eta}{D}}(\Lambda(t) + D)\right], t \ge 0.$$

Due to unavoidable imperfections in the control of manufacturing processes, lifetimes of items in a population are probabilistically heterogeneous (a heterogeneous population). Generally, a heterogeneous population consists of two subpopulations, i.e., a weak subpopulation and a normal subpopulation. Compared with the items in the normal subpopulation, items in the weak subpopulation have a shorter mean lifetime and are prone to give rise to early in-use failures. For example, it is widely believed that integrated circuits consist of a small proportion of weak items with much shorter lifetimes than the normal items. To eliminate weak items, engineers place each and every item under elevated temperature or voltage for a certain period of time and release those items which survive the test to field service. This test is known as (traditional) burn-in test. Traditional burn-in tests that stress items to failure are inefficient for highly reliable products among which even the weak items will take a long time to fail. This predicament can be tactfully solved in the condition-based burn-in when there are some suitable quality characteristics. In the context of condition-based burn-in, all products are exercised and, at the end of the process, items with deterioration levels below the cut-off point will be released to field service whereas items with deterioration levels exceeding the cut-off point will be discarded. Because the cut-off point is usually much lower than the failure threshold, the condition-based burn-in can tactfully handle high reliable products.

The rest of this chapter is organized as follows. Section 2 gives the total cost functional. Two types of misspecification errors are introduced. Section 3 deals with the maximum likelihood estimation of the unknown parameters in the proposed degradation models. Section 4 uses a simulative example to illustrate the proposed method. Some concluding remarks are given at the end of this chapter.

2 Optimal Burn-in Policy

Consider a heterogeneous population with a small proportion of weak items. We assume that the deterioration of an item from the normal population has an inverse Gaussian process $IG\left(\Lambda_1(t), \eta\Lambda_1(t)^2\right)$, whereas the deterioration of an item from

the weak population has an inverse Gaussian process $IG\left(\Lambda_2(t), \eta\Lambda_2(t)^2\right)$, with $\Lambda_1(t) > \Lambda_1(t) > 0$, for all $t > 0$. Set $r(t)$ to be the cut-off point for $X(t)$ with the decision rule: An item is classified to be normal if and only if at time t the deterioration level is smaller than the cut-off point, i.e., $X(t) \le r(t)$.

For a fixed time t, the probability of misclassifying a normal item as a weak item (type-I error) is

$$\begin{aligned}\alpha(t) = P(X(t) > r(t)|\textit{normal item}) &= 1 - F_{IG}\left(r(t); \Lambda_1(t), \eta\Lambda_1(t)^2\right) \\ &= \Phi\left[\sqrt{\frac{\eta}{r(t)}}[\Lambda_1(t) - r(t)]\right] \\ &\quad + \exp\{2\eta\Lambda_1(t)\}\Phi\left[-\sqrt{\frac{\eta}{r(t)}}[\Lambda_1(t) + r(t)]\right], t \ge 0.\end{aligned}$$

The probability of misclassifying a weak item as a normal item (type-II error) is

$$\begin{aligned}\beta(t) = P(X(t) \le r(t)|\textit{weak item}) &= F_{IG}\left(r(t); \Lambda_2(t), \eta\Lambda_2(t)^2\right) \\ &= \Phi\left[\sqrt{\frac{\eta}{r(t)}}[r(t) - \Lambda_2(t)]\right] \\ &\quad + \exp\{2\eta\Lambda_2(t)\}\Phi\left[-\sqrt{\frac{\eta}{r(t)}}[\Lambda_2(t) + r(t)]\right], t \ge 0.\end{aligned}$$

Let n denote the number of items subject to a condition-based burn-in test and w the proportion of the weak items. The cost of misclassification is composed of the type-I cost and the type-II cost:

$$C_{misc}(t, r(t)) = n\left[C_\alpha \times (1 - w)\alpha(t) + C_\beta \times w\beta(t)\right].$$

Here, $C_\alpha(C_\beta)$ denotes the per-unit cost of misclassifying a normal (weak) item as a weak (normal) item. At time t, set $r^*(t)$ to denote the optimal cut-off point minimizing the cost of misclassification, i.e., $r^*(t) = \arg\min_{r(t) > 0} C_{misc}(t, r(t))$.

In the burn-in procedure, degradation measurements of an item are collected at epochs $t = 0, t_1, \ldots, t_l$ with each inspection costing C_{mea}. Hence, the number of inspections at time t_b is $b + 1$ for $1 \le b \le l$. The per-item cost of conducting the burn-in test per unit of time is denoted by C_{op}. Hence, the overall cost of conducting a burn-in test on n items up to time t_b is

$$TC(t_b, r(t_b)) = C_{misc}(t_b, r(t_b)) + nC_{op}t_b + nC_{mea}(b + 1).$$

With the available degradation measurements increasing, the cost of misclassification will decrease over time, whereas the cost of inspection and operation will increase over time. Hence, the optimal burn-in time should make a trade-off

between $C_{misc}(t_b, r(t_b))$ and $nC_{op}t_b + nC_{mea}(b+1)$. For each checking point t_b, the corresponding optimal cut-off point can be determined by minimizing the overall cost $TC(t_b, r(t_b))$. The globally optimal burn-in time t_b^* can then be determined by minimizing $TC(t_b, r^*(t_b))$, i.e. $t_b^* = \arg\min_{\{t_b\}_{b=1}^{l}} TC(t_b, r^*(t_b))$.

Note that the bivariate optimization problem,$\min_{b,r} TC(t_b, r(t_b))$, is an integer optimization problem in b and is equivalent to the optimization problem $\min_b \min_r TC(t_b, r(t_b))$. Since b is an integer-valued variable, the latter optimization problem can be easily solved by solving two univariate minimization problems. Specifically, by fixing b, the optimization problem $\min_r TC(t_b, r(t_b))$ can be solved by analytical or numerical methods; the resulted minimized cost is hence a function of b. By choosing b to be $1, 2, 3, \ldots$, we can find t_1^*, t_2^*, t_3^*,… respectively such that the corresponding $TC(t_1, r(t_1)), TC(t_2, r(t_2)), TC(t_3, r(t_3)), \ldots$ are minimized.

3 Estimating Unknown Parameters

The distribution of $X(t)$ is a mixture of two inverse Gaussian distributions. The probability density function is given by

$$P(X(t) = x) = (1-w) f_{IG}\left(x; \Lambda_1(t), \eta\Lambda_1(t)^2\right) + w f_{IG}\left(x; \Lambda_2(t), \eta\Lambda_2(t)^2\right),\ x \geq 0.$$

For all $1 \leq i \leq n$ and $1 \leq j \leq l$, we might denote the degradation increment of the ith item at time t_j as $x_{i,j} = X_i(t_j) - X_i(t_{j-1})$ with $t_0 = 0$. The increment $x_{i,j}$ is an observation from the mixture distribution $(1-w) f_{IG}\left(x; \Delta\Lambda_1(t_j), \eta\left[\Delta\Lambda_1(t_j)\right]^2\right) + w f_{IG}\left(x; \Delta\Lambda_2(t_j), \eta\left[\Delta\Lambda_2(t_j)\right]^2\right)$, where $\Delta\Lambda_1(t_j) = \Lambda_1(t_j) - \Lambda_1(t_{j-1})$, and $\Delta\Lambda_2(t_j) = \Lambda_2(t_j) - \Lambda_2(t_{j-1})$. Given the measured increments $\{x_{i,j}, 1 \leq i \leq n, 1 \leq j \leq l\}$, the resulted log-likelihood function is given by

$$l(\Lambda_1(t), \Lambda_2(t), \eta, w) = \sum_{i=1}^{n} \log\left\{(1-w)H_i^1 + wH_i^2\right\},$$

in which we have

$$H_i^k = \prod_{j=1}^{l} \sqrt{\frac{\eta}{2\pi x_{i,j}^3}} \times \Delta\Lambda_k(t_j) \times \exp\left\{-\frac{\eta\left[x_{i,j} - \Delta\Lambda_k(t_j)\right]^2}{2x_{i,j}}\right\}, k = 1, 2.$$

Once the functional forms of $\Lambda_1(t)$ and $\Lambda_2(t)$ are specified, we can obtain the maximum likelihood estimates on the unknown parameters using numerical methods. The optimal burn-in time and the optimal cut-off point can thereupon be obtained to weed out the weak items. The Akaike information criterion (AIC) can

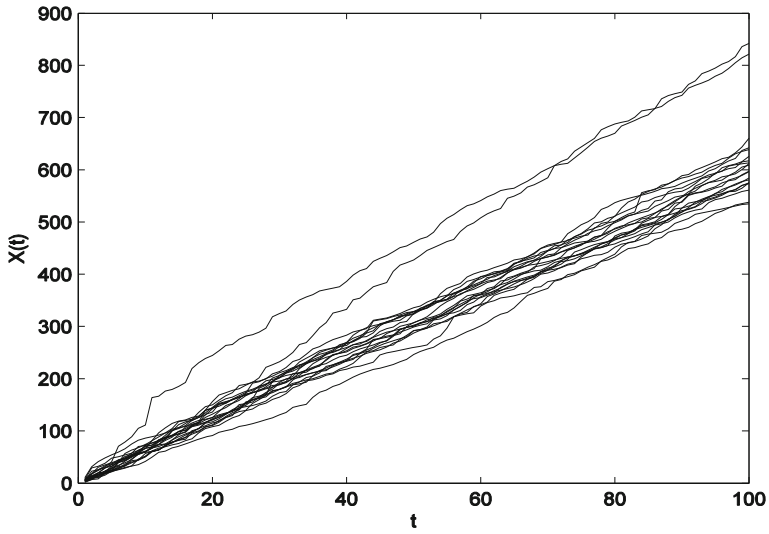


Fig. 1 Inverse Gaussian degradation paths with 20 entities

be served for testing goodness of fit, which is defined by $AIC = 2m - 2l$. Here, m is the number of model parameters and l is the maximized value of the log-likelihood function (Fig. 1).

4 A Numerical Example

A commonly used formulation of the marginal distribution in the inverse Gaussian process is given by $IG(ut^q, \eta t^{2q})$, and when $q = 1$ it reduces to a stationary stochastic process. For illustrative purpose, we use in this section the following data set: $u_1 = 0.0345, u_2 = 0.0519, \eta = 0.0521$, and $q = 1$. We consider here the unit time scale in hours. The measurements of the deterioration are made every 5 h. The proportion of the weak components is $w = 0.2646$. The cost configurations of the proposed burn-in test are given as: $C_\alpha = 65, C_\beta = 90, C_{op} = 0.09$, and $C_{mea} = 0.05$. Some degradation trajectories of the items from this heterogeneous population, with sample size being 20, are plotted in Fig. 2. As can be visualized, two components are nonconforming, degrading faster than the others.

At each checking point, we use a pattern search optimization method to grabble the corresponding optimal cut-off point, minimizing the total cost function. The optimal cut-off points and the related minimized costs, as well as the misclassifying probabilities, are listed in Table 1. From Table 1, we observe that the globally minimized total cost is 8.5811 at time 55 h. Hence, the globally optimal burn-in time turns out to be 55 h, and the corresponding globally optimal cut-off point is 2.3367. From Table 1, we can see that as the number of inspections increases, the

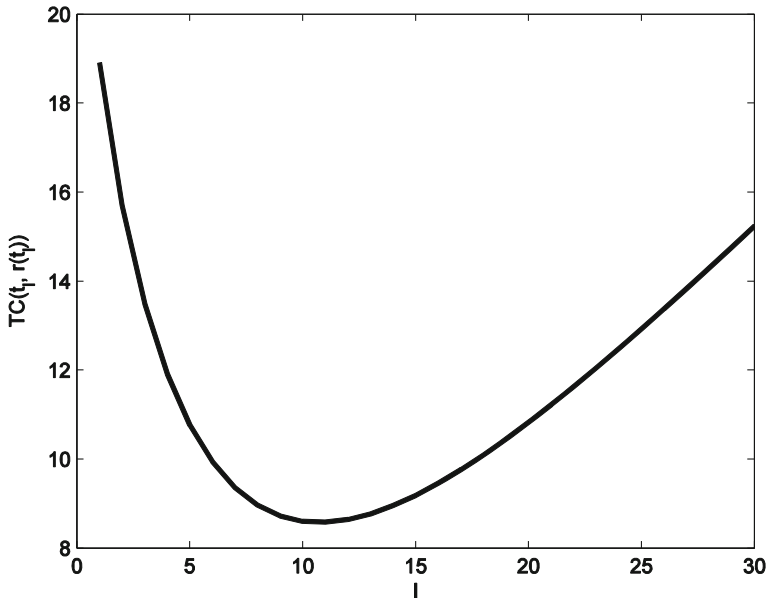


Fig. 2 The plot of minimized total cost versus the number of checking times

misclassifying probabilities, i.e. the probabilities of type-I and type-II errors, decrease. As a result, the cost of misspecification decreases. However, as the burn-in duration increases, the cost of operation increases. We plot the minimized total cost versus the number of checking times in Fig. 2. As can be seen, the minimized total cost firstly decreases and then increases, and the globally minimal total cost uniquely exists.

5 Conclusion

For highly reliable products, it is rather difficult to determine the optimal burn-in time due to the scarce of the event data. This problem can be solved successfully if there is a quality characteristic of which the degradation can be related to the reliability of the product. In this chapter, we presented a decision rule for classifying an item as a weak or a normal unit. The optimal cut-off point can only be obtained using numerical methods. However, most generally used optimization algorithms can be used to find the optimal cut-off point, as well as the minimized total cost. We use a numerical example to show that the optimal policy exists and is unique.

The stationary inverse Gaussian processes have a finite number of jumps in finite time intervals, and hence are suitable for modeling usage such as damage due to sporadic shocks. The stationary Gamma processes have an infinite number of jumps in finite time intervals, and hence are suitable for describing gradual damage by

Table 1 Optimal cut-off point, probabilities of Type-I and Type-II errors and minimized total cost at each inspection

l	1	2	3	4	5	6	7	8	9	10
$r^*(t)$	0.2643	0.4715	0.6787	0.8860	1.0933	1.3005	1.5078	1.7149	1.9222	2.1295
$\alpha(t)$	0.0845	0.0877	0.0799	0.0703	0.0611	0.0528	0.0455	0.0392	0.0338	0.0291
$\beta(t)$	0.6013	0.4391	0.3410	0.2729	0.2223	0.1833	0.1525	0.1276	0.1074	0.0907
$TC(t)$	18.9085	15.6993	13.4901	11.9096	10.7646	9.9390	9.3564	8.9637	8.7220	8.6019
l	11	12	13	14	15	16	17	18	19	20
$r^*(t)$	2.3367	2.5439	2.7512	2.9584	3.1656	3.3729	3.5801	3.7874	3.9946	4.2018
$\alpha(t)$	0.0251	0.0216	0.0187	0.0161	0.0139	0.0121	0.0104	0.0090	0.0078	0.0068
$\beta(t)$	0.0769	0.0654	0.0557	0.0476	0.0407	0.0349	0.0300	0.0258	0.0222	0.0191
$TC(t)$	8.5811	8.6418	8.7701	8.9547	9.1863	9.4574	9.7619	10.0946	10.4512	10.8281
l	21	22	23	24	25	26	27	28	29	30
$r^*(t)$	4.4091	4.6163	4.8236	5.0308	5.2380	5.4453	5.6525	5.8598	6.0670	6.2742
$\alpha(t)$	0.0059	0.0051	0.0044	0.0038	0.0033	0.0029	0.0025	0.0022	0.0019	0.0017
$\beta(t)$	0.0165	0.0142	0.0123	0.0106	0.0092	0.0079	0.0069	0.0060	0.0052	0.0045
$TC(t)$	11.2224	11.6314	12.0531	12.4856	12.9273	13.3771	13.8336	14.2960	14.7635	15.2354

continuous use. Since the inverse Gaussian process and the Gamma process both have monotone, discontinuous paths, they are competitive candidates for degradation-modeling problems, especially when the deterioration-checking points are discretely distributed. Therefore, the topic of misspecification, i.e., mis-specifying the stationary inverse Gaussian process (stationary Gamma process) as the stationary Gamma process (stationary inverse Gaussian process), when using maximum likelihood (ML) methods for estimation and inference, is of interest.

References

1. Wasan MT (1968) On an inverse Gaussian process. Skandinavisk Aktuarietidskrift 1968:69–96
2. Wang X, Xu D (2010) An inverse gaussian process model for degradation data. Technometrics 52:188–197
3. Ye ZS, Chen N. (2013) The inverse Gaussian process as a degradation model. Technometrics, to appear
4. Chikkara RS, Folks JL (1989) The inverse Gaussian distribution. Marcell Dekker, New York

Condition Based Maintenance and Operation of Wind Turbines

Tieling Zhang, Richard Dwight and Khaled El-Akruti

Abstract With application of advanced sensing technology, the condition based maintenance and operation has been made possible to many industrial systems. In a wind turbine, there are a few hundreds of sensing signals used to monitor the component performance and operational condition. The condition information is utilized in operational control of wind turbines and the wind farm in order to reduce the down time and Cost of Energy (CoE). In this chapter, a framework of condition based maintenance and operation of wind turbines is presented. This framework starts with data collection of sensing signals through SCADA and includes data processing and modeling, failure pattern recognition, remaining useful life/health condition prediction, load prediction (prediction of wind trend), integrated decision making for maintenance and operation of wind turbines and the wind farm, and maintenance planning. The research challenges involved in each step of the framework are discussed. The framework presented in this chapter serves as a guideline which is also useful to other systems.

1 Introduction

Wind power is one of the main clean and renewable energy sources. Its penetration in energy market keeps increasing in the past 20 years. The US Department of Energy aims to achieve 20 % of wind energy penetration in the utility market by the end of 2030 [1]. However, wind power contributed only about 3.5 % of the total

T. Zhang (✉) · R. Dwight · K. El-Akruti
School of Mechanical, Materials and Mechatronic Engineering, University of Wollongong, Wollongong, NSW 2522, Australia
e-mail: tieling@uow.edu.au

R. Dwight
e-mail: radwight@uow.edu.au

K. El-Akruti
e-mail: khaled@uow.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_87

electricity generated in the US in 2012 [2]. For the European Wind Energy Association, it is reported that the goal is to generate 26–34 % of the electricity from wind by 2030 [3]. And, China's wind industry is forecasted to reach 150 GW of installed capacity by 2015 which is well beyond the central government's goal of 100 GW by 2015 [4] and 230 GW of installed capacity by 2020 [5]. It is no doubt that the global market of wind energy is steadily growing.

Wind turbines are complex electromechanical systems usually having a design lifetime of 20–30 years. Wind turbine system reliability is a critical factor in the success of a wind energy project [6]. Studies have shown that the spending on wind turbine maintenance and repair accounts for 25–30 % of the life cycle cost (e.g., [7]). These have provided strong impetus for improvement on wind turbine reliability and optimization in maintenance and operation for reducing cost of energy (CoE).

Maintenance optimization is a crucial issue for industries that utilize physical assets due to its impact on costs, risks and performance [8]. In modern industry, the maintenance strategies have so far changed from the old-aged corrective and preventive one into the condition based maintenance (CBM) due to innovation and developments of sensing technology. Numerous applications can be found in today's industry. Condition monitoring based maintenance is that a maintenance service is scheduled based on the health status of a component/subsystem under monitoring. It needs a condition monitoring system implemented, for example, vibration measurements for essential mechanical components in wind turbines. The purpose of condition monitoring is to ensure continual operation of wind turbines with continuous measuring and analysis, and thereby increasing the turbine availability and reducing expenses. Especially in connection with offshore wind farms, detailed planning of maintenance based on the state of the turbines is an important requirement.

Condition based maintenance includes data acquisition through a condition monitoring system, data processing and modeling for health condition assessment and prediction, and decision making on service action. The techniques about CBM have been well developed and research in this area grows very fast. Hundreds of research papers in this area, including theoretical development and practical applications appear in every year [9], see, e.g., the review papers [9–11].

It is, however, until recent years that the CBM technologies have been implemented onto wind turbines in order to realize optimization in maintenance and operation. An early overview of condition monitoring techniques for wind turbines can be referred to [12]. This is a state-of-the-art review on the techniques up to 2002. Later on with the rapid increase of demand orders on large wind turbines, the wind turbine technology has been enhanced very fast and the maintenance optimization has become an issue to notice in wind farm's operation. Since more and more sensors are installed into modern wind turbines and more sensing signals are recorded, it makes the CBM become a real practice and the main focus of maintenance strategies of the wind farm operators. Maintenance optimization with condition monitoring has then become a hot research topic and received much attention. Garcia et al. [13] proposed an intelligent system for predictive

maintenance with application to the health condition monitoring of a wind turbine gearbox. Andrawus et al. [8] discussed the quantitative maintenance optimization for wind turbines using Monte Carlo simulation and Delay-Time Maintenance Model (DTMM). Nilsson and Bertling [14] discussed maintenance management of wind turbine systems using condition monitoring systems by focusing on life cycle cost analysis. Lu et al. [15] conducted a state-of-the-art review on condition monitoring and fault diagnosis for wind turbines and point out that "although many techniques existing in other industries can be directly or indirectly applied, wind turbines present particular challenges for successful and reliable diagnostics and prognostics". Gray and Watson [16] discussed an approach based on physics of failure to wind turbine CBM. Byon and Ding [17] developed models and the associated solution tools for devising optimal maintenance strategies. They consider a multi-state deterioration model for wind turbines subject to different failure modes. Besnard and Bertling [18] developed an approach for condition-based maintenance optimization applied to wind turbine blades. This approach is applicable to inspection-based maintenance as well as online condition-monitoring based maintenance. Nielsen and Sørensen [19] studied risk-based operation and maintenance of offshore wind turbine components by considering lifetime costs related to all the activities in life cycle. Tian et al. [20, 21] developed a CBM policy to address the maintenance optimization at the wind farm level where the key component health condition is predicted by neural networks. Amayri et al. [22] proposed a CBM method by considering different types of wind turbines in one wind farm. Van Horenbeek et al. [23] studied a prognostic maintenance policy which makes use of predictive information, i.e., the remaining useful life of different components of the wind turbine. This policy is applied to the whole wind farm rather than one wind turbine with considerations of the importance of dependencies between separate wind turbines in a wind farm, especially an offshore wind farm. Dong, et al. [24] proposed a systematic multi-parameter health condition evaluation framework that considers the dynamic operational environment of wind turbines. Shafiee [25] developed an opportunistic condition-based maintenance strategy for offshore wind turbine blade under cold weather conditions. With the maintenance strategy, an opportunistic maintenance action is performed for non-failed blades while to execute replacement of the damaged ones of offshore wind turbines.

As summarized above, CBM systems and technologies have been developed and more advanced techniques with practical applications will be seen in the near future. In order to realize the optimal operation of wind turbines in service life, however, it is not enough just depending on execution of CBM with advanced technologies developed because of uncertainty of wind load. Wind cannot be forecasted with higher accuracy in a few days ahead. This makes it become a big challenge to realize and maintain optimization in maintenance and operation of wind turbines and wind farm. The intermittency of wind power renders its specialty from other industrial sectors where work load can be predicted. It is therefore not only CBM but also CBO (condition based operation) are needed in realizing the optimal operation of wind turbines in service life. Although wind turbine suppliers, and wind farm operators and owners have noticed the importance of CBM and

CBO of wind farm, the dedicated and mature research results have not been found to be reported. Erickson and Sauers [26] described the concept of intelligent wind turbines in a project conducted at Los Alamos National Laboratory (USA) and led by a multi-disciplinary team of experts covering structural health monitoring, modeling and simulation and prognostic decision-making. The Intelligent Wind Turbines team is developing predictive models, advanced sensing technologies, novel data interrogation techniques, active performance control and reliability-based decision-making algorithms to address important issues that currently hinder the wind industry.

Researchers and wind turbine manufacturers have started to make efforts towards development and implementation of CBM/CBO technologies in wind farm operational management. It is envisaged that new methods, techniques and systems will appear in not far future for condition based operation of wind turbines and wind farm. This is the motivation of this present chapter. The purpose of this chapter is to present a framework for realizing CBM/CBO of a wind farm. The framework serves as a guideline for establishment of CBO systems for a wind farm. A few technical schemes associated with CBM/CBO for extending remaining useful lifetime (RUL) of turbine components are also discussed.

2 A Framework for CBM/CBO

In this section, we present a framework for CBM/CBO of wind turbines. This framework is shown in Fig. 1. Condition monitoring systems include Condition Monitoring System & Solution (CMS&S) and Turbine Monitor (TM). Many

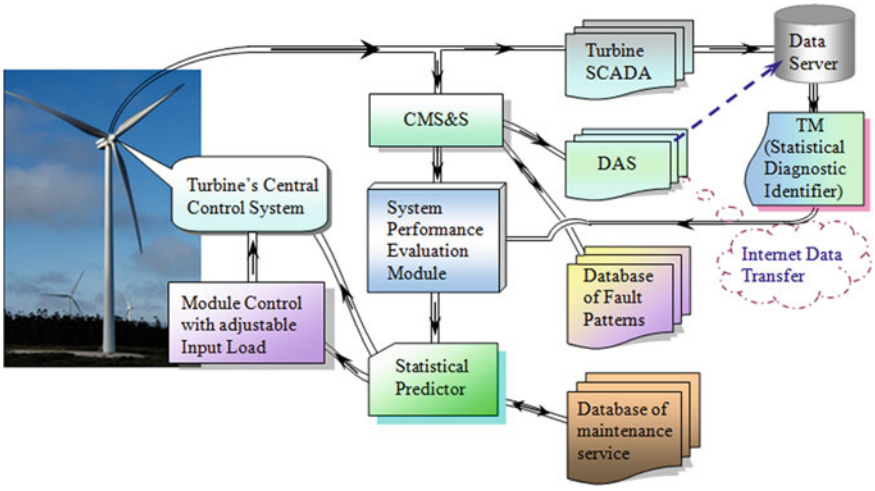


Fig. 1 Framework for CBM/CBO of wind turbines

sensors are implemented into wind turbines for condition monitoring and control. The sensor signal data are transferred to CMS&S and turbine SCADA. All sensor signal raw data may be stored in turbine SCADA. CMS&S is a function module that monitors and reports the components' and subsystems' health status by focusing on failure mechanisms using turbine control and operation signals such as phase voltage and current, power production, etc. as well as other signals measured directly from components and subsystems such as vibration, acoustic, RF signals and temperature.

After data processing in CMS&S, the output signal data are transferred to Data Acquisition System (DAS) where the historical data of sensor signals and model output are stored. On the other hand, the component and subsystem health status information are stored in DAS and DAS has another function to communicate with other turbines' DAS in the same wind farm or communicate with Farm SCADA through the internet for data transmission. In addition, the Data Server serving for wind turbine monitors (TM) may request data from DAS in addition to its request from Turbine SCADA. Data Server stores part of Turbine SCADA signals. Not all signal data stored in Turbine SCADA are stored in Data Server. The data stored in Data Server are the signals which may be used often or the data representing the typical characteristics of turbines.

TM is based on historical data recorded. One turbine monitor is a statistical model which can be a distribution model, or a regression formula (linear or non-linear), or some other kinds of models derived from historical data. The outputs from TM are transferred to System Performance Evaluation Module for system level, subsystem level and component level performance evaluation.

Database of Fault Patterns serves to store all possible failure patterns and severity of failure modes that are obtained in design phase, function test and reliability test phase, historical failure cases as well as those obtained from analyzing FMEA and Fault Tree Analysis (FTA). CMS&S may continuously generate new failure patterns and keep updating of Database of Fault Patterns. Database of Fault Patterns can be a stand along one or may be imbedded into CMS&S as one function.

Different level alarms issued by CMS&S and TM are transmitted to System Performance Evaluation Module for subsystem or system performance evaluation. If a subsystem is evaluated as in a wear-out process, the severity of performance degradation will be presented. This information is used as a reference for Turbine Central Control System to take action if an adjustable input load can be applied. Such an adjustable input load can be a reduced load or enhanced load to a subsystem or turbine. The alarms and turbine health status are also sent to turbine operator and turbine maintenance technicians.

The System Performance Evaluation Module consists at least of one wind turbine performance evaluation index system, an algorithm module (Program of built models), and a tool module including hardware and interface.

- The wind turbine performance evaluation index system includes a standard for evaluation of subsystem performances: What are the factors selected in the performance evaluation system and how to determine the levels for each of the factors.
- The algorithm module may use Data mining techniques and other expert system methods to find out the relation between one failure mode and its relevant impact factors. The relations are used for performance degradation evaluation. It involves a method to predict performance degradation either for component or for subsystems.
- A tool module is a realization of the above two aspects. It includes hardware structure and interfaces, etc.

All the information passed through System Performance Evaluation Module for turbine system performance evaluation will be transferred to Statistical Predictor for further modeling and analysis.

Statistical Predictor includes different functional modules to complete modeling of time to failure (TTF) data, reliability and availability modeling and analysis. It can include the following functions:

- Reliability and availability models with multiple system states where the transition rate between different states can be a function of time.
- Function of Monte Carlo simulation for system reliability and availability solution.
- Function of predicting the next maintenance time.

The output from this module is a clear solution that is if the turbine can keep running until the next scheduled periodical maintenance time under normal control and operation. If it is not, can the turbine operate with reduced load so that the turbine can keep operation without shut-down till the next periodical maintenance? Or, if not, what is the maintenance plan? The outputs of Statistical Predictor also include a report with maintenance plan, time used in maintenance service, method of maintenance, number of spare parts, maintenance cost, etc.

If no alarms appear both from CMS&S and TM, or it is verified that the turbine can continuously run under normal condition without shut-down until the next periodical maintenance time, the adjustable load control scheme is not initiated. Otherwise, an adjustable load control plan is initiated and applied to turbine central control.

Database of Maintenance Service stores historical data and report data from Statistical Predictor, typical maintenance cases with number of spare parts used, maintenance cost, maintenance time distribution, number of spare parts in inventory, etc. CBM/CBO based on condition monitoring may depend only on CMS&S or TM.

Currently, each large wind turbine has been equipped with condition monitoring and fault diagnostic system. Figure 2 shows a flow diagram of data processing and modeling process for CBM/CBO. The models built are most of linear correlation model. The linear correlation between one variable and other related factors is

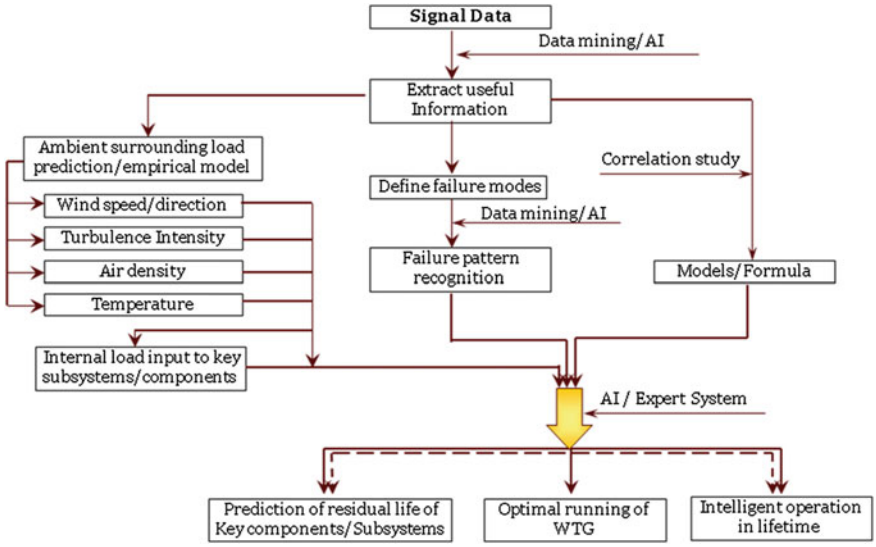


Fig. 2 Schematic of data processing and modeling process for CBM/CBO

hardly maintained because of complexity of the wind turbine system and dynamic load environment. It is, therefore, quite often to explore the nonlinear models for a failure mode or a degradation case study. To build such models needs one's experience and the pre-existing examples. Another way to handle the problem is by using Matlab System Identification Module. This module can help build the non-linear model without the need of previous experience and knowledge. One the other hand, one needs to look at the over-fitting of a model built and the false alarm rate in the verification. The trade-off study is often required before a good model is found and finally put into application.

The models developed, using condition monitoring information, are used for diagnostics and prognostics of components and subsystems. It is also important to know the remaining useful lifetime of the critical components and subsystems in order to maintain the turbine's operation until the next scheduled maintenance service with a planned control scheme. In the next section, a few technical schemes for extending RUL are discussed in CBO for wind turbines.

3 Condition Based Maintenance/Operation of Wind Turbines

It needs to carry out advanced research for realization of condition based operation (CBO) of a wind farm. The aim of CBO is to enable better informed operational decision-making, resulting in optimal power production by leveraging CBM

oriented information, reduction of unscheduled shutdown of wind turbines, increase of availability and hence Mean Time between Inspections (MTBI). The first step to run CBO is to get clear information about component health status and then derive or estimate the remaining useful lifetime under a given operational condition. If it is found that a component or some of components are in a severe status, it needs to take actions in operation either to reduce load in order to keep the turbine running without failure or it needs an immediate plan to conduct a maintenance service. The second step is to estimate/forecast the wind load in a few months ahead. CBO extends the remaining lifetime of components and enables the economic optimization of the entire production process based on wind of on-site prediction. Wind cannot be predicted with higher accuracy even it is just in a few days ahead but the high and low wind season is clearly known for an area such as weekly, biweekly, or monthly average of the wind can be plotted in history. It is expected that all the maintenance services are scheduled to be performed in the low wind season. Therefore, it is important and needed to extend the remaining useful life of the components whose performance (health status) has degraded.

3.1 Scheme for Extending Remaining Useful Lifetime

Here, suppose a component vibration level keeps increasing due to some failure mode. A typical one of such components is gearbox or generator bearing. Figure 3 shows an illustration of RUL after one alert level is triggered. Each curve marked with different power output class represents that the turbine is operating in that power output class with more than 80 % of the remaining useful lifetime. Here the

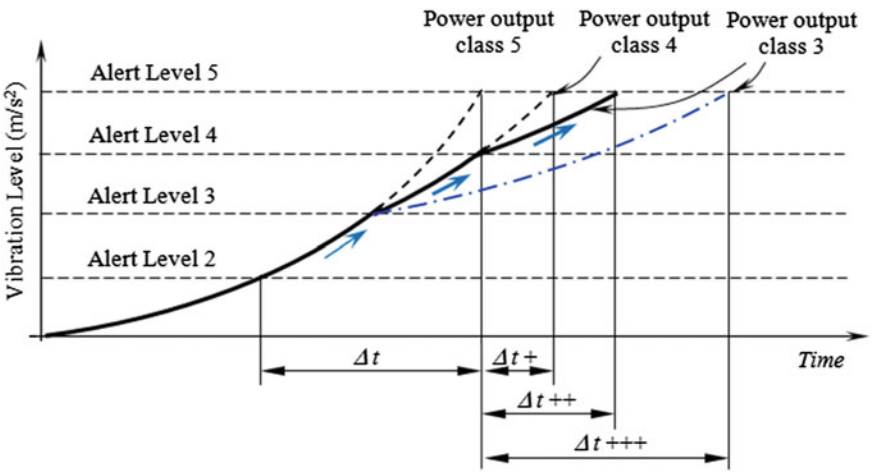


Fig. 3 Illustration of a turbine's operation with extended remaining useful lifetime

Table 1 Definition of power output class for 2.0 MW WTG

Power output class	1	2	3	4	5
Power value P_w (MW)	$P_w \leq 0.7$	$0.7 < P_w \leq 1.2$	$1.2 < P_w \leq 1.5$	$1.5 < P_w \leq 1.8$	$1.8 < P_w$

remaining useful lifetime means the time after Alert Level 2 is triggered till it reaches Alert Level 5, for instance.

Alert Level 5 is a severity level at which the component is required to be replaced in a very short time, say, one or two weeks. In another meaning, the component approaches the end of service life if Alert Level 5 is triggered. After Alert Level 2 is triggered, the remaining useful lifetime is Δt if the turbine is running with power output class 5. If this Δt is shorter than the scheduled time to do maintenance or the time to lower wind season, the turbine is controlled to run under power output class 4 after Alert Level 3 is triggered and then the remaining useful lifetime is extended as shown with $\Delta t+$. If the total time of Δt and $\Delta t+$ are still shorter than the scheduled time to do maintenance or time to lower wind season, the turbine may run under power output class 3 if a further alert level is triggered, see Fig. 3. Then, a further extension is obtained. That is, the extended RUL is represented by $\Delta t++$. An alternative operation is that the turbine may be controlled to run power output class 3 after Alert Level 3 is triggered, hence, a more longer extension of RUL is obtained, see $\Delta t+++$ in Fig. 3. The definition of different power output class is given in Table 1 for 2.0 MW WTG, for instance [27].

3.2 Approaches to Determination of RUL

As discussed above, RUL needs to be determined in CBM/CBO. In this section, we present three different ways to determine RUL of a turbine component. They are based on historical data, laboratory tests and regression analysis.

3.2.1 RUL Based on Historical Data

According to the operational history of the identical or similar wind turbines, the empirically obtained data may comprise data relating to lifetimes of identical/similar components. Such data can be obtained in the following manner. When a given alarm level is detected for a given component, this component is monitored and the power output as a function of time, and the time elapsing between the alarm level was detected until the component broke down is logged. This can be done for a large number of identical/similar components, thereby τ_i (i = 1–5) for each component case is recorded, where τ_i is the cumulative time elapsed under power output class i (i = 1–5, e.g.) from the designated alarm level detected to the time when the component fails. Then, the statistical information about RUL for a component under

various operating conditions is obtained such as RUL = $a_0 + a_1\tau_1 + \ldots + a_5\tau_5$, or other forms from regression analysis. Then, the expected RUL in the future can be calculated with scheduled and controlled operating conditions.

3.2.2 RUL Obtained from Lab Tests

Another approach to obtaining RUL is by laboratory testing. In the test, a severity level of component can be set up and then run the test with a planned scheme of load levels corresponding to the power output class 1–5. Through a number of tests, the RUL distribution corresponding to each power output class level can be obtained. These distributions will be then used for risk assessment for an expected RUL with a designed scheme associated with designated operating conditions.

3.2.3 RUL Based on Regression Analysis with Multiple Inputs

Since a number of sensing signals have been recorded from each wind turbine, one can carry out regression analyses of the selected variables for a specific problem. The regression equation is then established based on the historical data recorded. After this regression equation is verified, it will be used to estimate RUL of a component by a manner to calculate the distance of means between a group distribution and a distribution from each wind turbine. This distance can be evaluated by 1σ, 2σ, 3σ, etc. representing different severity levels. Here, σ is standard deviation of the group distribution. The expected RUL can be obtained by adjusting the variable values in turbine's control of operation. This way has been utilized to build wind turbine monitor as the one which is the same as Vestas Turbine Monitor (VTM). Vestas is a leader supplier of wind turbines in global wind energy market.

4 Conclusion

In this chapter, a framework for CBM/CBO is presented, which is composed of condition monitoring system, data acquisition module, data processing and modeling module, system performance evaluation index module and so on. This framework serves as a guideline when considering development of CBM/CBO systems. With innovation and development of new sensors, and advanced diagnostic and prognostic technology, it can be seen that the optimal operation of wind turbines based on condition monitoring will come to practice in not far future. From quality and reliability point of view, the benefits brought about by CBO can be summarized as follows:

- Prevent premature shutdown of wind turbine. Early fault detection will prevent catastrophic failures from occurrence during high wind season through de-rating control operation. Thus, this extends the service lifetimes of key components and subsystems, e.g., gearbox and generator of a wind turbine without compromising the wind turbine and overall wind farm performance.
- Help improve capacity factor, loss of production factor, etc. When a good estimation of the RUL is available, repairs or replacement action can be scheduled during time frames with very low wind. Capacity factor is defined as the ratio of the actual production over a given period of time to the amount of power the turbine would have produced running at full capacity during the same time period.
- Support for further development. CMS provides detailed information on the dynamic behavior of a wind turbine over a long period of time. The increased knowledge about the system can be used for supporting the new design of wind turbine components.
- Increase system reliability. Because the component health status is monitored, the turbine operation is controlled according to the severity of components and predicted wind on site. The unscheduled shutdown times of wind turbines will be reduced especially when a turbine is running in a high wind season.

The big challenge to run CBO, however, is wind load prediction with required accuracy. In addition, the challenges may also come from other aspects, for example, the followings:

- WTGs with CMS are in their earlier and mid-career phase. Thus only a few failed cases are available so that there is no enough data to calculate RUL.
- The models' confidence level is higher with increased number of historical data verified. However, it needs the data logged in a longer time.
- The CBO efficiency with de-rating control for extending RUL may depend on the failure mode type.
- The de-rating control of turbines cannot help realize the optimal power output.
- The methodology presented in this chapter requires real life data. However, due to short time data and availability, test-rig is of a good choice to support. But, carrying out lab tests still take a lot of time and the test condition may not well emulate the turbine's operating condition.

Therefore, there exist a lot of research challenges in front of us. It requires us to make efforts to find solutions for those issues. The further research concerning CBO applications will include but not least the followings:

- Development of CBO schemes
- Other approaches to determination of RUL
- De-rating control techniques for CBO
- System architectures for realization of CBO for a wind farm

- New approaches and techniques for CBO optimization by considering power tariff and other constraints
- Development of a requirements analysis framework for CBM/CBO of offshore wind farm [28].

References

1. The US Department of Energy (US DoE) 20 % wind energy by 2030: Increasing wind energy's contribution to US electricity supply. http://www1.eere.energy.gov/wind/wind_energy_report.html
2. Kenward A (2013) Forecast dims for future growth in wind power. http://www.climatecentral.org/news/forecast-dims-for-future-growth-in-wind-power-15721. Accessed on 12 Mar 2013
3. European Wind Energy Association (2009) Pure Power – Wind energy targets for 2020 and 2030, a report by the European Wind Energy Association—2009 update. http://www.ewea.org/fileadmin/ewea_documents/documents/publications/reports/Pure_Power_Full_Report.pdf
4. Marcacci S (2012) China forecast to hit 150 GW installed wind capacity by 2015. http://cleantechnica.com/2012/11/30/china-forecast-to-hit-150-gw-installed-wind-capacity-by-2015
5. China's wind power forecast at 230 GW by 2020 (2010) The World's #1 Renewable Energy Network for News & Information. http://www.renewableenergyworld.com/rea/news/article/2010/10/chinas-wind-power-forecast-at-230-gw-by-2020
6. Walford CA (2006) Wind turbine reliability: understanding and minimizing wind turbine operation and maintenance costs. Sandia report SAND.2006-1100, Sandia National Laboratories
7. Yang W, Tavner PJ, Wilkinson MR (2009) Condition monitoring and fault diagnosis of a wind turbine synchronous generator drive train. IET Renew Power Gener 3(1):1–11
8. Andrawus JA, Watson J, Kishk M (2007) Wind turbine maintenance optimization: principles of quantitative maintenance optimization. Wind Eng 31(2):101–110
9. Jardine AKS, Lin DM, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 20(7):1483–1510
10. Peng Y, Dong M, Zuo MJ (2010) Current status of machine prognostics in condition-based maintenance: a review. Int J Adv Manufact Technol 50(1–4):297–313
11. Sikorska JZ, Hodkiewicz M, Ma L (2011) Prognostic modeling options for remaining useful life estimation by industry. Mech Syst Signal Process 25:1803–1836
12. Verbruggen TW (2003) Wind turbine operation AND maintenance based on condition monitoring. Final report in Apr 2003. ftp://ecn.nl/pub/www/library/report/2003/c03047.pdf
13. Garcia MC, Sanz-Bobi MA, del Pico J (2006) SIMAP: Intelligent System for Predictive Maintenance: Application to the health condition monitoring of a wind turbine gearbox. Comput Ind 57:552–568
14. Nilsson J, Bertling L (2007) Maintenance management of wind power systems using condition monitoring systems - life cycle cost analysis for two case studies. IEEE Trans Energy Convers 22:223–229
15. Lu B, Li YY, Yang ZZ (2009) A review of recent advances in wind turbine condition monitoring and fault diagnosis. PEMWA 2009–power electronics and machines in wind applications, 2009. 24–26 June 2009, Lincoln, NE, USA
16. Gray CS, Watson SJ (2009) Physics of failure approach to wind turbine condition based maintenance. Wind Energy 13(5):395–405
17. Byon E, Ding Y (2010) Season-dependent condition-based maintenance for a wind turbine using a partially observed Markov decision process. IEEE Trans Power Syst 25(4):1823–1834

18. Besnard F, Bertling L (2010) An approach for condition-based maintenance optimization applied to wind turbine blades. IEEE Trans. Sustain Energy 1(2):77–83
19. Nielsen JJ, Sørensen JD (2011) On risk-based operation and maintenance of offshore wind turbine components. Reliab Eng Syst Saf 96:218–229
20. Tian ZG, Jin TD, Wu BR, Ding FF (2010) Condition based maintenance optimization for wind power generation systems under continuous monitoring. Renew Energy 36:1502–1509
21. Tian ZG, Ding Y, Ding FF (2011) Maintenance optimization of wind turbine systems based on intelligent prediction tools. Innovative Computing Methods and Their Applications to Engineering Problems Studies in Computational Intelligence 357:53–71
22. Amayri A, Tian Z G, Jin T D (2011) Condition based maintenance of wind turbine systems considering different turbine types. In: International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering ICQR2MSE, 17–19 June 2011, Xi'an, pp 596–600
23. Van Horenbeek A, Van Ostaeyen J, Duflou J, Pintelon L (2012) Prognostic maintenance scheduling for offshore wind turbine farms. In: 4th world conference production and operations management, Amsterdam, The Netherlands, 1–5 July 2012
24. Dong YL, Fang F, Gu YJ (2013) Dynamic evaluation of wind turbine health condition based on Gaussian mixture model and evidential reasoning. J Renew Sustain Energy 5(3).http://dx.doi.org/10.1063/1.4808018
25. Shafiee M (2013) An opportunistic condition-based maintenance strategy for offshore wind turbine blade under cold weather conditions. In: International wind energy conference on winterwind, 12–13 Feb 2013, Östersund, Sweden
26. Erickson M, Sauers A (2011) Making wind turbines intelligent. Innovation - America's J Technol Commercialization 9(3):27
27. Zhang T, Zhou Y, Lim K P, et al (2013) Method and system for controlling operation of a wind turbine. United States Patent, Patent No. US 8355823 B2, 15 Jan 2013
28. El-Thalji I, Jantunen E (2012) On the development of condition based maintenance strategy for offshore wind farm: requirement elicitation process. Energy Procedia 24:328–339

Status of Using, Manufacturing and Testing of Ethylene Pyrolysis Furnace Tubes in China

T. Chen, X.D. Chen, Y.R. Lu, Z.B. Ai and Z.C. Fan

Abstract In this paper, the problems of short service life, main failure modes and failure mechanisms of ethylene pyrolysis furnace tubes in China were summed up through service condition survey on ethylene pyrolysis furnace tubes in the domestic nine petrochemical corporations, compared with the situation of two ethylene enterprises in the developed countries. Chemical composition, tensile properties at room temperature and high-temperature stress rupture properties of furnace tubes and fittings were obtained through experiments performed on furnace tubes and fittings from six furnace tube manufacturing enterprises. The revision suggestion of centrifugally cast alloy tubes standard and the relationship between newly manufactured furnace tubes and fittings performance and serviced furnace tube failure modes were also discussed in present article.

Keywords Ethylene pyrolysis furnace tubes · Status of using · Manufacture · Testing

1 Introduction

Centrifugally cast heat resistant alloys composed of high chromium and nickel content are widely used as material of ethylene pyrolysis furnace tube, the service temperature is generally in the range of 900–1,150 °C [1–4]. According to API 530, the design life of furnace tube is 10^5 h (11.4 a). A survey aiming at obtaining the status of ethylene pyrolysis furnace tubes was performed by Sinopec since 2008. The results show that the service life of ethylene pyrolysis furnace tubes is mostly 3–5 a and accidental failures often occur, which baffles the long-period operation in petrochemical corporatins [5]. However, the service life of foreign furnace tubes can generally be employed for 6–8 a continuously without shutdown unplanned. In comparison,

T. Chen (✉) · X.D. Chen · Y.R. Lu · Z.B. Ai · Z.C. Fan
Hefei General Machinery Research Institute, Hefei 230031, People's Republic of China
e-mail: chentao@hgmri.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_88

service life of domestic ethylene pyrolysis furnace tubes is much shorter. "Sinopec Inspection and Assessment Center on Furnace Tube, SIACFT" was set up in Hefei General Machinery Research Institute to carry out inspection and assessment of newly manufacturing, and in-service furnace tubes in order to strengthen the management of manufacturing quality of the furnace tube, to improve the service life, and to provide guarantee for the long period operation of ethylene pyrolysis installation.

At present, the standards relating to ethylene pyrolysis furnace tubes include HG/T 2601-2011 Centrifugal casting alloy tubes for service of pressure bearing at high temperature [6], HG/T 3673-2011 Static cast fittings of furnace for service pressure bearing at high temperature [7] and ASTM A608 Centrifugally cast Iron-Chromium-Nickel high-alloy tubing for pressure application at high temperatures [8], etc. The requirement of the centrifugally cast furnace tube and static casting fittings in the standard mentioned above is too loose, which does not adapt the development of petrochemical industry. Many problems have been found in the past two years by SIACFT. The status of using, manufacturing and testing of ethylene pyrolysis furnace tubes in China is summed up in this paper in order to enhance the understanding of furnace tube and then provide guarantee for the long period operation of ethylene pyrolysis installation.

2 Using Status of Ethylene Pyrolysis Furnace Tube in Petrochemical Corporation

Sinopec entrusted Hefei General Machinery Research Institute (SIACFT) to carry out the investigation of radiant furnace tubes of ethylene pyrolysis furnace in order to understand the using status and analyze the cause of early failure of furnace tube. Till July 2011, nine Sinopec ethylene corporations provide account and using status of ethylene pyrolysis furnace tube to SIACFT.

The maximum designed ethylene output of 9 Sinopec corporations in the survey is 9.26 million tons per year, and the designed treatment capacity of a single furnace ranges from 15 thousand tons per year to 0.23 million tons per year. The related furnace type is complex, such as SL-I, SL-II, SL-IV, SRT-III, SRT-IV, CBL-I, CBL-III, CBL-IV, CBL-R, USC, GK-V, GK-VI, KTI-GKV, etc. Figure 1 shows statistics of replaced furnace tube in different years of 9 corporations after ethylene pyrolysis furnace put into operation. The number of units which were totally replaced is 14 in less than 5 years, accounting for about 14 %; in 6–8 years, the number is 44, which is the largest, accounting for about 44 %; in 9–10 years, the number is 22; more than 10 years, the number is 22.

Take Company A for example, Fig. 2 shows the using states of ethylene pyrolysis furnace tubes. It shows that replacement of furnace tubes and fittings is mainly caused by 6 factors, such as design of the furnace, creep, medium, quality of fitting, carburization and stress. After CBL-III furnace put into operation for 2–3 years, 128 furnace tubes were replaced totally due to tube bending caused by the factor of the design of the furnace. Tubes of 6 furnaces relating to SRT-IV and SL-I

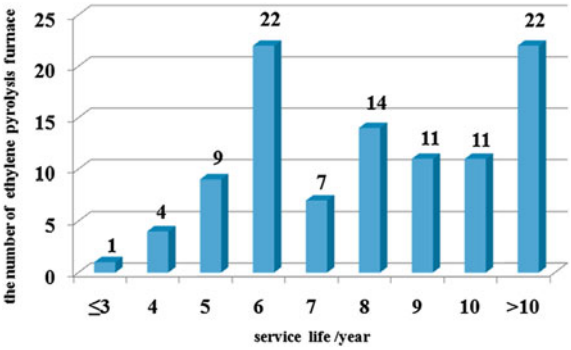


Fig. 1 The number of replacement of the whole ethylene pyrolysis furnace in different years

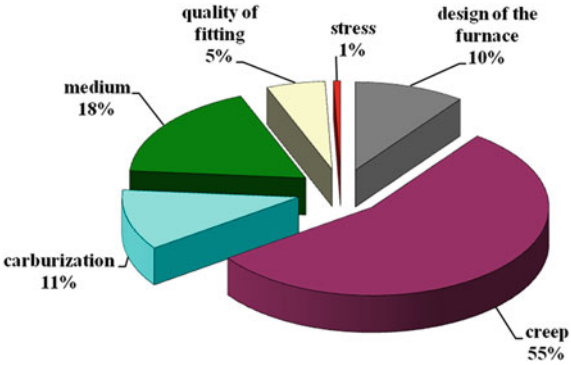


Fig. 2 Using status of ethylene pyrolysis furnace tube of company A

suffered from creep: several replacement of tubes occurred in 3 years after operation, 180 tubes replacement occurred in 5 years, 167 tubes in 5–8 years, and 348 tubes in 8–9 years. 140 tubes/fittings replacements occurred because of carburization relating to 6 sets of furnace in 6 years. 223 pieces of weld metal in 4 different sets of furnace suffered from corrosion because of medium factor in 2–3 years after operation. 68 fittings failure occurred because of poor quality involving 7 sets of furnace. The detailed records are given as following: the number of replaced fittings is 1, 3, 18, 32, 9, and 5, respectively in 1.5, 2, 3, 4, 5, 6 years after operation. 8 pieces of tubes are replaced because of the install stress factor in less than 5 years after operation. Thus, creep and carburization factors accounted for about 66 % are the most important factors resulting in the replacement of furnace tube of the enterprise.

Correspondingly, ethylene enterprises in German and Italy don't have extra ethylene pyrolysis furnaces as backup according to the survey performed by SI-ACFT, which is different from domestic ethylene pyrolysis corporations. In order to avoid shutdown accidentally, the furnace tubes are disposable integral replaced on time according to schedule, without considering the actual damage of tubes. An ethylene enterprise in German makes a one-time replacement of ethylene pyrolysis furnace tubes after putting into operation for 6 years. As for an ethylene enterprise in Italy, the using time of the exit section of furnace tube is 6 years and the using

time of the entrance section is 8 years. The ethylene pyrolysis furnaces of the two enterprises operate continuously, without any repair. Accident failure seldom takes place during operation. So it shows that foreign furnace tube product has good quality uniformity and stability. However, the replacement of domestic furnace tubes of ethylene pyrolysis furnace often occurred in 3–5 years after operation. Therefore, there is a obvious gap of the quality stability between foreign and domestic furnace tube, frequent replacements not only lead to economic loss and increase operating costs of enterprises, but also affect the long period operation.

3 Manufacturing of Ethylene Cracking Furnace Tube

Generally, the intermediate frequency furnace is chosen to melt raw material of furnace tube and centrifugal casting machine is selected to manufacture tubes. Figure 3 shows the typical melting and casting field of furnace tube. Due to the high service temperature of ethylene pyrolysis furnace tubes, it is clearly pointed out in the ordering technical specifications that waste furnace tube cannot be doped. Because the impurity removal ability of the melting process is quite weak, the requirement of the cleaning of the raw materials is very strict. Therefore, it focuses on the cleaning of the raw materials in order to ensure the cleaning of smelting material in the manufacturing process.

Fig. 3 Typical melting and casting field of furnace tube

There are about 10 centrifugal casting furnace tube manufacture plants in China. Raw materials are provided by domestic or foreign plants. Generally, the contents of impurities in raw materials including S, P, Pb, As, etc. are analyzed by optical emission spectrophotometry. Chemical composition of casting product in furnace tube manufacture plants; is also analyzed by optical emission spectrophotometry only one plant is equipped with inductive coupled plasma emission spectrometer. There are control index of impurities in each plant. For example, the index of P ≤ 0.02 wt.%, S ≤ 0.012 wt.%, Pb ≤ 50 ppm, Sn ≤ 0.01 wt.% is required in a casting plant. At the same time, management of raw materials in most pants is disordered according to the survey performed by SIACFT.

There are two famous furnace tube manufacturers in Europe. According to the investigation of company B in France and company C in German, the two manufactures don't have superiority to domestic manufacturers in chemical composition test. Optical emission spectrophotometry is also used to analyze chemical composition of furnace tube. But they have internal standard on raw materials selection. Specific value can't be provided because of the intellectual property right. Raw material is stacked neatly and strictly marked by nameplate. It is clearly pointed out that the waste furnace tube cannot be doped by the two foreign manufacture plants.

In addition, in order to obtain high-performance centrifugal casting ethylene pyrolysis furnace tube, centrifugal casting parameters should be considered during the casting process, such as inner wall coating of metal mould, casting temperature, the vibration amplitude of centrifugal casting machine, preheating of metal mould and cooling speed, etc.

4 Testing of Ethylene Pyrolysis Furnace Tube

25Cr35NiNb alloy and 35Cr45NiNb alloy are widely used as ethylene pyrolysis furnace tube material. For 25Cr35NiNb alloy, typical chemical composition is: C: 0.35–0.50 wt.%, Cr: 23–27 wt.%, Ni: 34–37 wt.%, Nb: 0.8–1.5 wt.%. For 35Cr45NiNb alloy, typical chemical composition is: C: 0.40–0.60 wt.%, Cr: 30–37 wt.%, Ni: 43–47 wt.%, Nb: 0.6–1.8 wt.%.

The requirements of centrifugal casting furnace tubes are fairly loose in HG/T 2601-2011 Centrifugal casting alloy tubes for service of pressure bearing at high temperature, which can not adapt to the current development of petrochemical industry. Some indexes of centrifugal casting furnace tubes are significantly higher than those of HG/T 2601-2011 by designers. The content of some harmful elements are required as follow: S ≤ 0.03 wt.%, P ≤ 0.03 wt.%, Pb ≤ 100 ppm, As ≤ 100 ppm, Sn ≤ 100 ppm. As for high temperature endurance performance, the rupture life of 25Cr35NiNb alloy is required to be atleast 120 h at 1,100 °C under a stress of 17 MPa. For 35Cr45NiNb alloy, the value is also atleast 120 h at 1,100 °C under a stress of 16 MPa.

But a large number of tests should be carried out to validate whether control index in HG/T 2601-2011 or technical specifications proposed by the designer is

reasonable or not. In this paper, 38 pieces of 25Cr35NiNb alloy and 47 pieces of 35Cr45NiNb alloy furnace tubes are selected to analyze the chemical compositions and stress rupture test, in order to statistically analyze the level of current domestic centrifugal casting furnace tube and to provide an important basis for revision of the standard.

4.1 25Cr35NiNb Alloy Furnace Tubes

38 pieces of centrifugal casting furnace tubes were chosen to analyze the chemical compositions and stress rupture test. Figure 4 shows chemical composition distribution map of 38 pieces of 25Cr35NiNb alloy furnace tubes. The statistics content of C, Si, Mn, S, P, Cr, Ni, Nb is listed as following: the content range of C is 0.35–0.489 wt.%, the percent of pass is 100 % relatively to standard value; the content range of Si is 1.51–2.0 wt.%, the percent of pass is 100 %; the content range of Mn is 0.53–1.47 wt.% the percent of pass is 100 %; the content range of P is 0.0093–0.022 wt.%, the percent of pass is 100 % relatively to standard value; the content range of S is 0.0068–0.02 wt.%, the percent of pass is 100 % relatively to standard value; the content range of Cr is 24.5–26.8 wt.%, the percent of pass is 100 % relatively to standard value; the content range of Ni is 33.92–37.81 wt.%, the percent of pass is 92.1 % relatively to standard value; the content range of Nb is 0.71–1.5 wt.%, the percent of pass is 89.47 % relatively to standard value.

Figure 5 shows the statistics measured content of As, Sn, Pb, Bi of 25Cr35NiNb alloy furnace tubes. The trace elements content range is listed as following: As: 2.925–41.21 ppm; Sn: 4.04–27.31 ppm; Pb: 1.90–40.074 ppm; Bi: 0.011–0.66 ppm.

Stress rupture life at 1,100 °C/17MPa of 25Cr35NiNb alloy furnace tubes is given in Fig. 6. The rupture life is in the range of 8–427.38 h, which pass rate is only 82.50 % compared to standard value of 120 h. There is still a gap between test data and technical specifications.

4.2 35Cr45NiNb Alloy Furnace Tubes

Figure 7 shows chemical composition distribution map of 47 pieces of 35Cr45NiNb alloy furnace tubes. The statistics content of C, Si, Mn, S, P, Cr, Ni, Nb is listed as following: the content range of C is 0.401–0.60 wt.%, the percent of pass is 100 % relatively to standard value; the content range of Si is 1.31–1.80 wt.%, the percent of pass is 100 % relatively to standard value; the content range of Mn is 0.6–1.29 wt.% the percent of pass is 100 % relatively to standard value; the content range of P is 0.01–0.026 wt.%, the percent of pass is 100 % relatively to standard value; the content range of S is 0.0062–0.012 wt.%, the percent of pass is 100 %

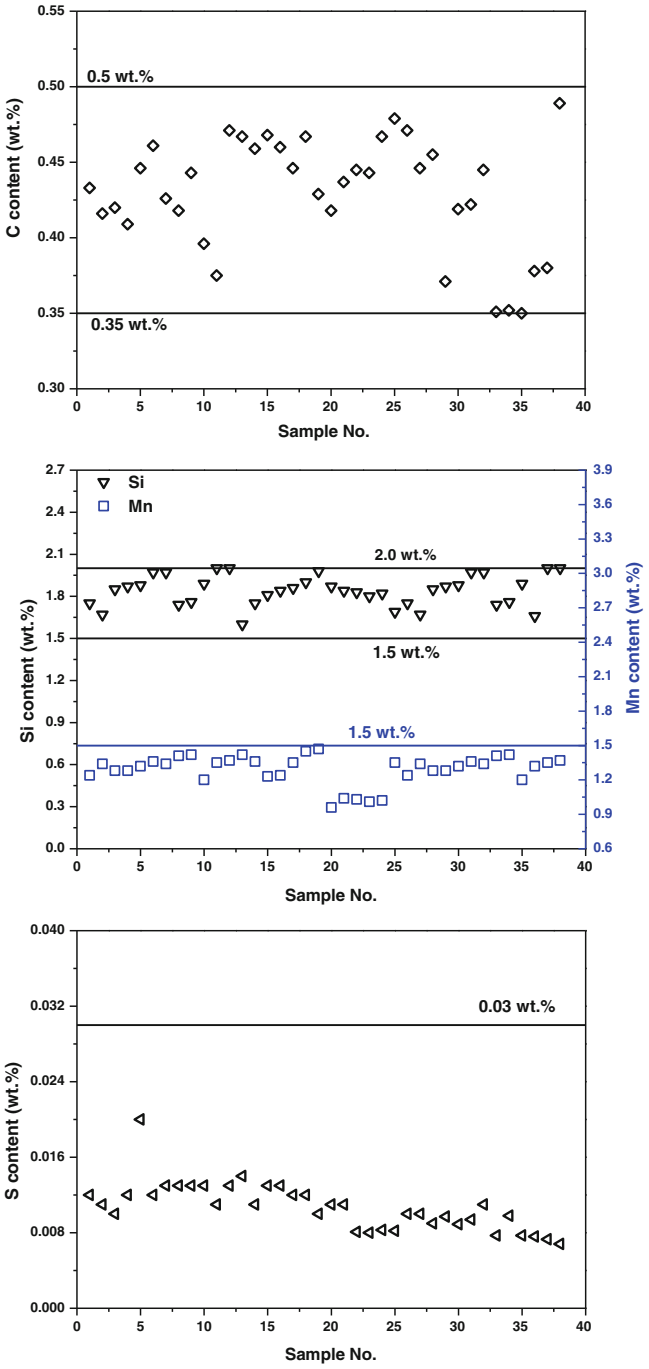


Fig. 4 Chemical composition distribution map of 25Cr35NiNb alloy furnace tubes

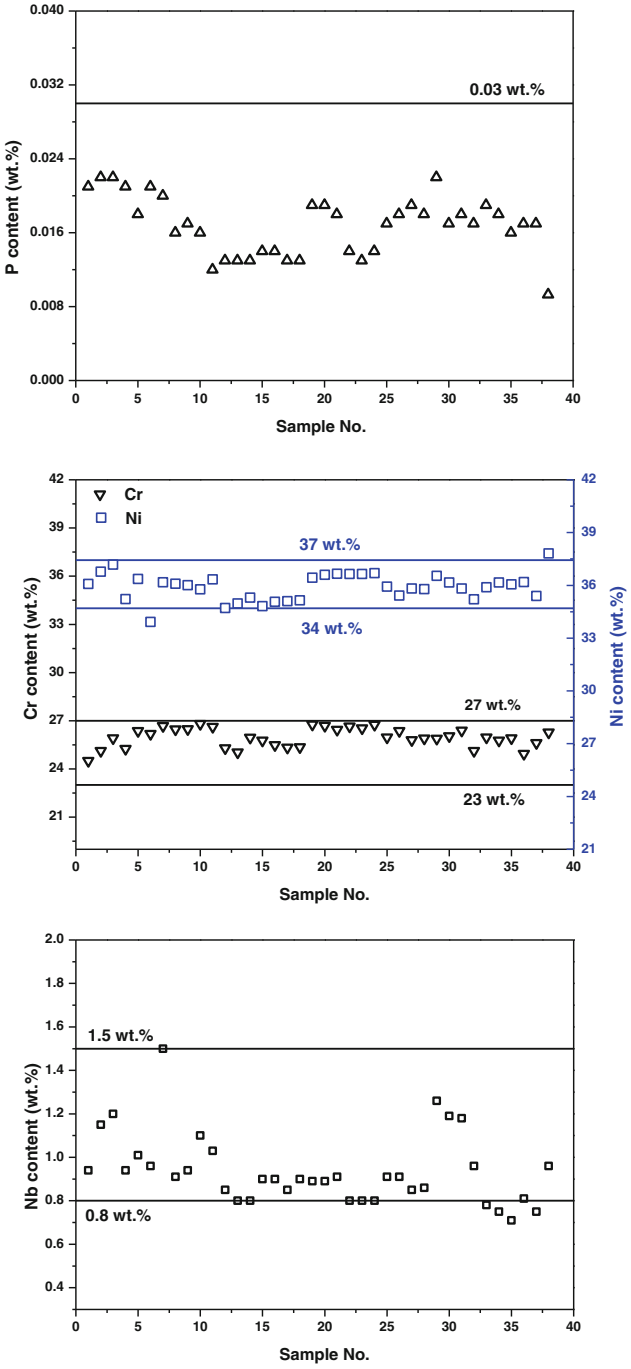


Fig. 4 continued

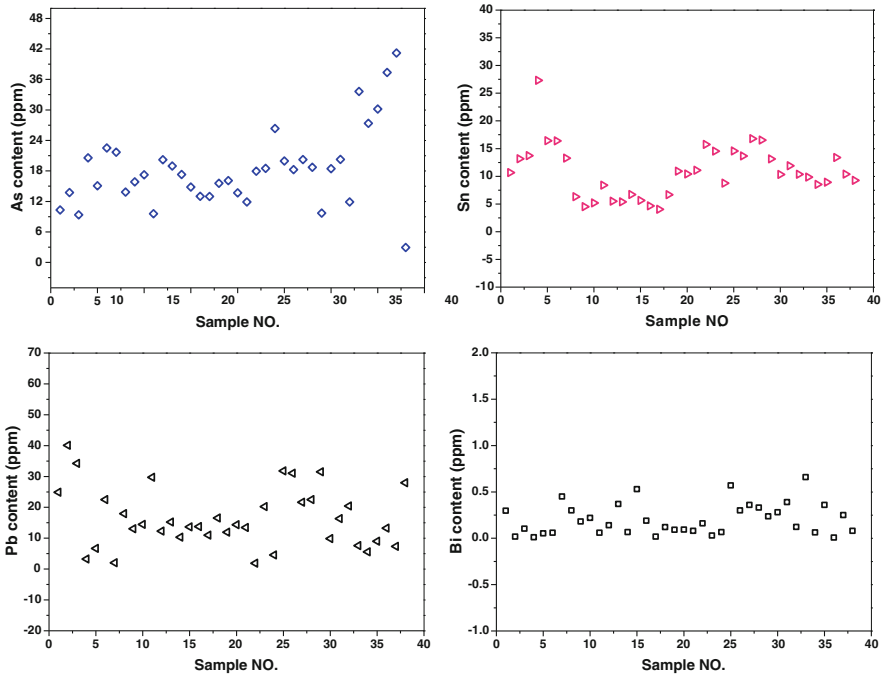


Fig. 5 The statistics measured content of trace elements of 25Cr35NiNb alloy furnace tubes

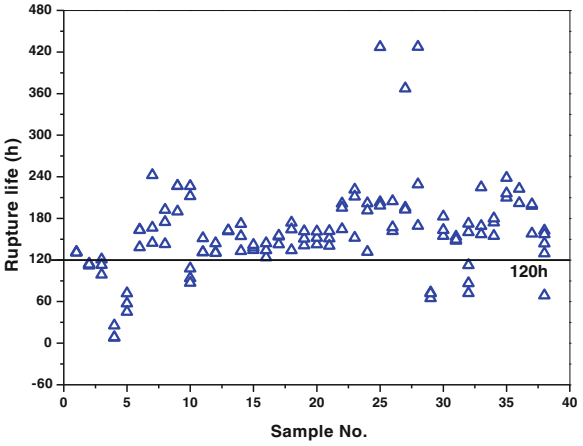


Fig. 6 Stress rupture life of 25Cr35NiNb alloy furnace tubes

relatively to standard value; the content range of Cr is 32.78–35.42 wt.%, the percent of pass is 100 % relatively to standard value; the content range of Ni is 42.77–47.02 wt.%, the percent of pass is 92.1 % relatively to standard value; the content range of Nb is 0.67–1.2 wt.%, the percent of pass is 89.36 % relatively to

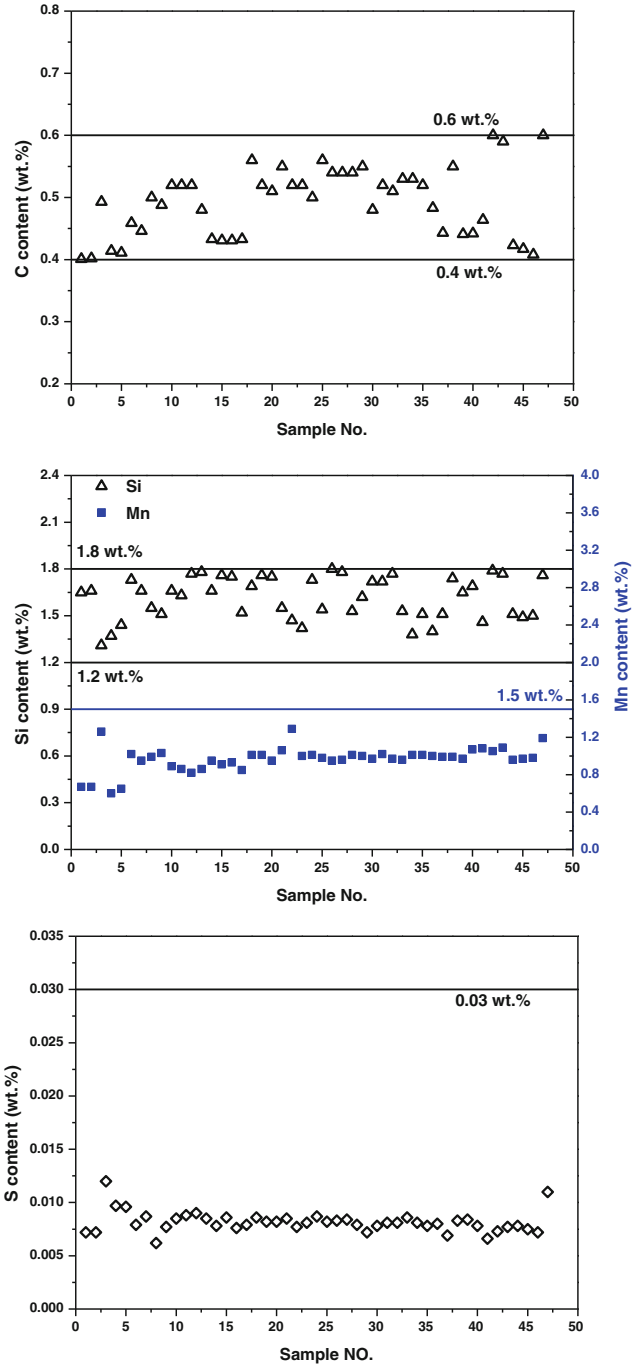


Fig. 7 Chemical composition distribution map of 35Cr45NiNb alloy furnace tube

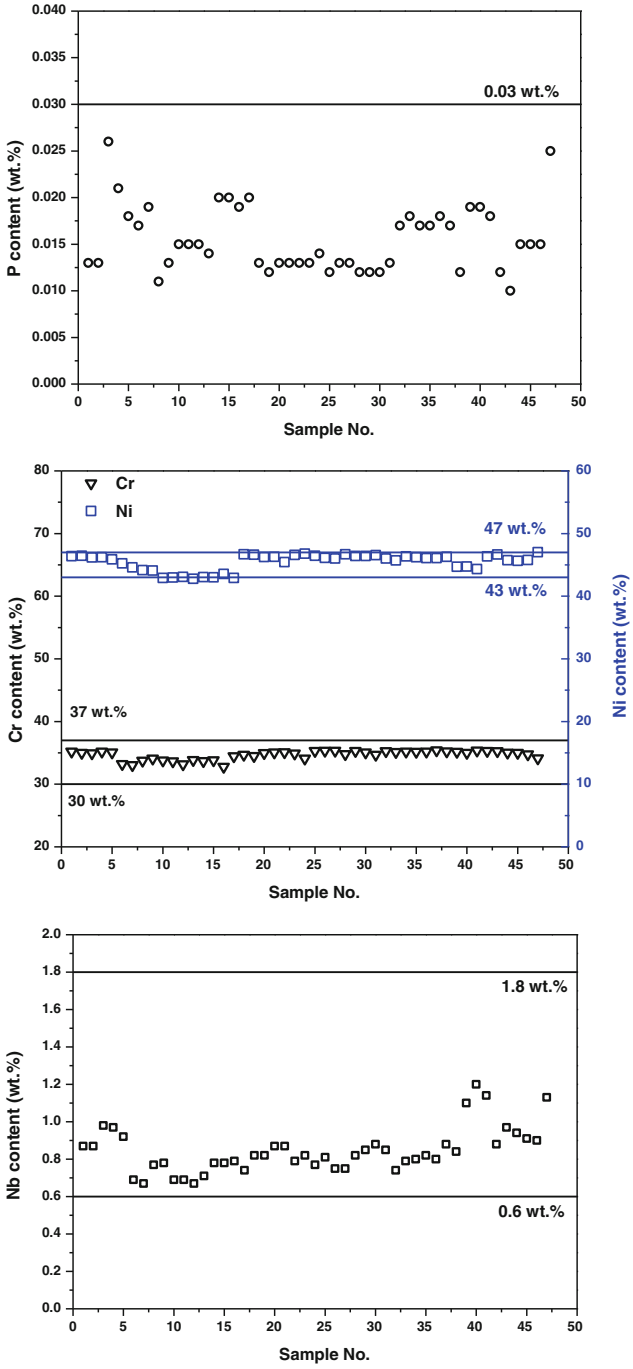


Fig. 7 continued

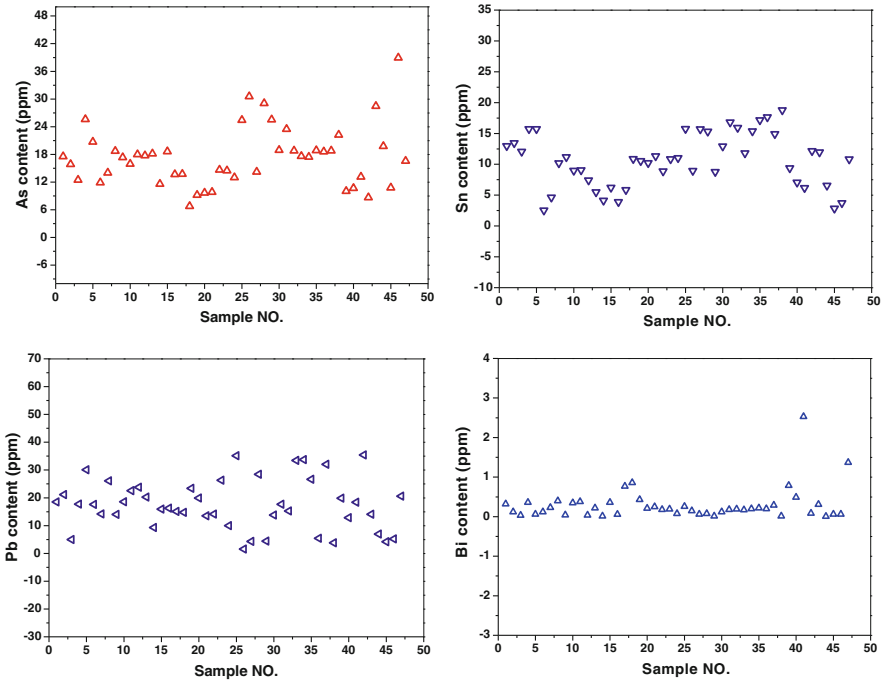


Fig. 8 The statistics measured content of trace elements of 35Cr45NiNb alloy furnace tubes

standard value. As to 35Cr45NiNb alloy furnace tubes, the content of P is lower than 0.026 wt.%, and the content of S is lower than 0.012 wt.%.

Figure 8 shows the statistics measured content of As, Sn, Pb, Bi of 35Cr45NiNb alloy furnace tubes. The trace elements content range is listed as following: As: 6.77–38.96 ppm; Sn: 2.52–18.77 ppm; Pb: 1.57–35.14 ppm; Bi: 0.005–2.53 ppm.

Stress rupture life at 1,100 °C/16 MPa of 35Cr45NiNb alloy furnace tubes is given in Fig. 9. The rupture life is in the range of 10.21–280.40 h, whose pass rate is only 83.22 % compared to standard value of 120 h. There is still a gap between test data and technical condition.

5 Discussions

5.1 Threshold Value of Impurity

Impurities such as S, Pb, Bi, As, Sn, etc. have attracted much attention on the effect of creep and fracture. A lot of researches have been carried out on pure nickel iron, low alloy steel, and nickel based alloy [9–13]. However, little work has been performed on the effect of impurities such as S, Pb, Bi, etc. on ethylene pyrolysis

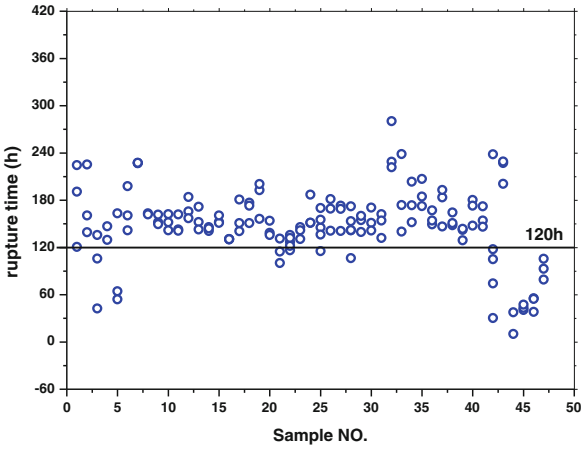


Fig. 9 Stress rupture life of 35Cr45NiNb alloy furnace tubes

furnace tube material, i.e., centrifugal casting heat resistant alloy composed of high chromium and nickel content. For centrifugal casting heat-resistant alloy furnace tubes, the requirement of the content of S should be lower than 0.030 wt.% in HG/T 2601 and 0.040 wt.% in ASTM A608. Therefore, the requirements of the content of S in domestic and foreign standards of centrifugal casting furnace tube are relatively low compared to the value obtained in this paper. There is no restriction on Pb and Bi in HG/T 2601 and ASTM A608. The item is generally proposed in HG/T 2601 that all of ethylene pyrolysis furnace tubes should be made of pure raw materials without adopting waste furnace tubes. There is no basis in foreign technical specifications requirements that the content of Pb, Bi and As should be no larger than 100 ppm compared to the value obtained in this paper.

Unfortunately, the mechanism of impurities on high temperature creep resistance of ethylene pyrolysis furnace tube material is not clear. The threshold value of impurity cannot be accurately proposed till now. So preliminary recommendations for 25Cr35NiNb alloy and 35Cr45NiNb alloy are listed as following: P ≤ 0.020 wt.%, S ≤ 0.015 wt.%, As ≤ 40 ppm, Sn ≤ 30 ppm, Pb ≤ 40 ppm, Bi ≤ 5 ppm according to the tests by SIACFT.

However, researches on effect of impurities on creep resistance of the furnace tube are still needed to be carried out and the relationship between impurities and rupture life should be established to determine the threshold of impurities more accurately.

5.2 Comparison Between Failure Mode and Quality of Furnace Tube

Creep and carburization are two main failure factors of domestic ethylene pyrolysis furnace tube according to survey in millions of tons of several ethylene corporations.

Fig. 10 Creep bending and drum kits of furnace tube

As for the creep factor, related to the creep elongation/bending, drum kits (as shown in Fig. 10), and creep rupture, the current technical specifications is that high temperature creep rupture life should not be less than 120 h at 1,100 °C under a stress of 16/17 MPa. Through the detection on stress rupture life of 25Cr35NiNb alloy and 35Cr45NiNb alloy materials, the pass rate of rupture life is 80–90 %, which cannot entirely accord with the requirement of technical condition. Creep performance evaluation is suggested to be considered in technical indicators of new manufacturing furnace tube.

The inner wall of ethylene pyrolysis furnace tube in use consistent may be suffered by carburizing (as shown in Fig. 11). Domestic researches are with those in America and Japan, that showing the carburizing accounted for 30–40 %. The carburized layer produces additional stress in the tube and makes changes on microstructure and properties of furnace tube, even can cause cracking of ethylene pyrolysis furnace tube. Similarly, anti-carbonation performance parameter in some test conditions has not been provided in the technical specifications. Anti-carbonation performance evaluation is suggested to be considered in technical indicators of new manufacturing furnace tube.

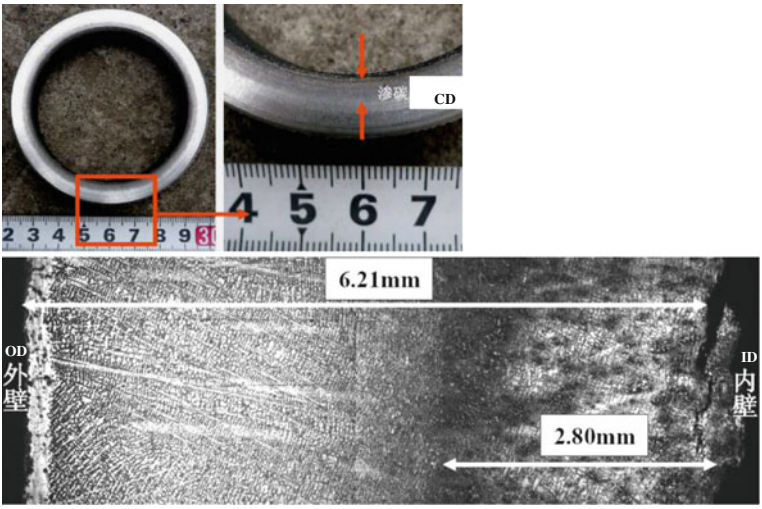


Fig. 11 Carburizing of furnace tube

6 Conclusions

(1) A lot of domestic ethylene pyrolysis furnace tubes have suffered replacement in 2–5 years due to creep elongation/bending, fracture and carburizing. Compared with few replacement records of foreign ethylene pyrolysis furnace tube in 8 years, there is an obvious gap between the quality of domestic and foreign furnace tube.

(2) For 25Cr35NiNb and 35Cr45NiNb alloy materials, the preliminary suggestions for the chemical content is listed as following: P ≤ 0.020 wt.%, S ≤ 0.015 wt.%, As ≤ 40 ppm, Sn ≤ 30 ppm, Pb ≤ 40 ppm, Bi ≤ 5 ppm.

(3) For 25Cr35NiNb and 35Cr45NiNb alloy furnace tubes, the pass rate of rupture life is 82–84 %, which affected the long period operation of ethylene pyrolysis furnace tube.

Acknowledgments This work was financially supported by the international science and technology cooperation project (2010DFB42960), Key Projects in the National Science & Technology Pillar Program during the 12th Five-Year Plan Period (2012BAK13B03), Science and Technology Fund of Excellent Youth of Anhui Province (Grand No. 1308085JGD04), and the Foundation for Scientific Research Institute from the Ministry of Science and Technology of China (Grand No. 2011EG219117).

References

1. Guan KS, Xu H, Wang ZW (2005) Quantitative study of creep cavity area of HP40 furnace tubes. Nucl Eng Des 235:1447–1456
2. Liu CJ, Chen Y (2011) Variations of the microstructure and mechanical properties of HP40Nb hydrogen reformer tube with time at elevated temperature. Mater Des 32:2507–2512
3. Branzaa T, Deschaux-Beaumeb F, Sierrab G, Loursa P (2009) Study and prevention of cracking during weld-repair of heat-resistant cast steels. J Mater Process Technol 209:536–547
4. Swaminathan J, Guguloth K, Gunjan MK, Roy P, Ghosh R (2008) Failure analysis and remaining life assessment of service exposed primary reformer heater tubes. Eng Fail Anal 15:311–331
5. Chen T, Chen XD, Lu YR, Lian XM, Ye J (2012) Influence of grain shape on rupture life of centrifugal casting 25Cr35Ni-Nb alloy tubes. In: Proceeding of ICPVT-13, London, UK
6. HG/T 2601-2011 Centrifugal casting alloy tubes service of pressure bearing at high temperature
7. HG/T 3673-2011 Static cast fittings of furnace for service pressure bearing at high temperature
8. ASTM A608 Standard specification for centrifugally cast iron-chromium-nickel high-alloy tubing for pressure application at high temperatures
9. George EP, Kennedy RL, Pope DP (1998) Review of trace element effects on high-temperature fracture of Fe- and Ni-base alloys. Phys Stat Sol 167:313–333
10. White CL, Schneibel JH, Padgett RA (1983) High temperature embrittlement of Ni and Ni-Cr alloys by trace elements. Metall Trans 14:595–610
11. George EP, Pope DP, Sklenicka V (1992) Clean steel technology. ASM International, Materials Park, p 17
12. Thomas GB, Gibbons TB (1984) Creep and fracture of a cast Ni-Cr-base alloy containing trace elements. Mater Sci Eng 67:13–23
13. Osgerby S, Gibbons TB (1984) The effect of trace elements on the creep behaviour of an Ni-Cr-base alloy. Mater Sci Eng A157:63–71

Improving Concrete Durability for Sewerage Applications

P.L. Ng and A.K.H. Kwan

Abstract Concrete is a widely adopted construction material in sewerage applications such as concrete pipes, manholes, box culverts, treatment tanks, and sewage conveyance tunnels. However, the contaminants in sewage may cause physical and chemical attacks to the concrete. In particular, the biogenic sulphuric acid attack poses a great threat to the concrete. This would shorten the service life of concrete and necessitate more frequent repairs and rehabilitations, thereby increasing the life-cycle cost of the sewerage infrastructure. As the prime solution to this problem, the authors advocate the development of *sewerage concrete* by improving the durability of concrete against sewerage attack. This chapter addresses the possible ways to improve the durability of concrete against sewerage attack, including the use of protective coatings, better concrete mix design (or more specifically mix design to improve the biogenic sulphuric acid resistance of concrete), and use of corrosion inhibitors.

1 Introduction

Concrete is a widely adopted construction material in sewerage applications including but not limited to concrete sewer pipes, manholes, box culverts, septic tanks, treatment tanks, and sewage conveyance tunnels (vitrified clay sewer pipes are also commonly used but concrete cannot be easily replaced for its ability to be cast into any desired geometry). Broadly speaking, urban sewage may be classified into domestic and industrial sewages. Domestic sewage encompasses residential and commercial effluents. It is typically composed of sulphide, chloride, ammonia and nitrogen compounds, and suspended solids [17]. Industrial sewage encompasses

P.L. Ng (✉) · A.K.H. Kwan
Department of Civil Engineering, The University of Hong Kong, Hong Kong, China
e-mail: irdngpl@gmail.com

A.K.H. Kwan
e-mail: khkwan@hku.hk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_89

effluents from various industrial processes. Its components and levels of contaminants vary significantly from industry to industry. All in all, the contaminants in sewage can cause various physical and chemical attacks to the concrete. In particular, as will be explicated in this chapter, the biogenic sulphuric acid attack poses a great threat to the concrete. Due to the corrosion effect of biogenic sulphuric acid, concrete is susceptible to deterioration and disintegration [16, 29, 33]. This would significantly shorten the service life of concrete and necessitate frequent repairs and rehabilitations, thereby increasing the maintenance cost and hence the life-cycle cost of the sewerage infrastructure.

As the prime solution to the above problem, the authors advocate the enhancement of durability of concrete for sewerage applications. The resistance of concrete against the corrosion effect of biogenic sulphuric acid should be particularly addressed. Many practitioners do not clearly distinguish sulphate attack from biogenic sulphuric acid attack, and there is a common misconception that improving the sulphate resistance of concrete will improve also the sulphuric acid resistance. Actually, it should be noted that biogenic sulphuric acid attack is an entirely different phenomenon from sulphate attack [24]. In the following, the authors will elucidate the mechanism of biogenic sulphuric acid attack in sewerage structures. The means of improving concrete durability for sewerage applications via the use of protective coatings, better concrete mix design, and use of corrosion inhibitors will be discussed.

Other factors affecting the durability of concrete in sewerage infrastructure include chloride attack from seawater flushing and from the aggressive underground environment, and carbonation of concrete promoted by the abundance of carbon dioxide from anaerobic microbiological decompositions [33]. These can be dealt with by better concrete mix design with the use of appropriate additives, use of high-performance concrete, use of protective coatings, and more careful crack control [15, 20]. The adoption of design standards of *marine concrete* to tackle chloride attack and carbonation is apposite [1, 10, 26]. In this respect, the measures to enhance the durability of concrete in marine environment are quite established. Hence, in the remaining of this chapter, the authors will address principally the issues regarding biogenic sulphuric acid attack. It is advocated herein to develop *sewerage concrete* for resisting biogenic sulphuric acid attack.

2 Biogenic Sulphuric Acid Attack of Concrete

Sewage contains abundant organic matters and sulphate. Besides, as the sewerage system is designed as enclosed system, the availability of oxygen is limited. Under this environment, anaerobic sulphate-reducing bacteria (SRB) such as *Desulfovibrio* and *Desulfobulbus* decompose the organic matter and reduce the sulphate to produce hydrogen sulphide (H_2S) gas as well as carbon dioxide (CO_2) gas [12]. The hydrogen sulphide is not destructive to the concrete, but it volatizes into the sewer atmosphere

(headspace) and dissolve in the slime layer (biofilm) coating the concrete surface above the liquid level. The carbon dioxide also volatizes into the headspace and combines with moisture to form carbonic acid (H_2CO_3), which promotes the eventual carbonation of concrete [33].

The hydrogen sulphide deposited in the biofilm at the headspace and crown undergoes oxidization by aerobic sulphur-oxidizing bacteria (SOB) to produce sulphuric acid (H_2SO_4). The SOB are typically belonging to *Thiobacilli* species, including *Thiobacillus thioparus*, *Thiobacillus novellus*, *Thiobacillus neapolitanus*, *Thiobacillus intermedius*, *Thiobacillus thiooxidans*, *Thiobacillus perometabolis*, and *Thiobacillus versutus* [21, 28, 32]. As the sulphuric acid is generated by a series of microbiological activities, it is referred to as biogenic sulphuric acid. The concentration of biogenic sulphuric acid is dependent on the type and population density of the SOB, temperature, and the flow regime of sewage that influence the relative abundance of oxygen. In many cases, the pH value can be lower than 2 [28]. For the biochemical aspects of SRB and SOB and their relation to sulphuric acid formation, readers may consult relevant literatures [12, 14].

The biogenic sulphuric acid attacks concrete in the following manner. At first, it reacts with the calcium silicate hydrates (gel product of cement hydration) and calcium hydroxide (lime) to produce gypsum ($CaSO_4$):

$$\mathrm{CaO \cdot SiO_2 \cdot 2H_2O + H_2SO_4 \rightarrow CaSO_4 + Si(OH)_4 + H_2O}$$
$$\mathrm{Ca(OH)_2 + H_2SO_4 \rightarrow CaSO_4 \cdot 2H_2O}$$

In the next step, the gypsum reacts with the hydrated tricalcium aluminates (C_3A) of cement to form ettringite ($2Ca_3Al(OH)_6 \cdot 3SO_4 \cdot 26H_2O$):

$$\mathrm{3CaO \cdot Al_2O_3 \cdot 12H_2O + 3(CaSO_4 \cdot 2H_2O) + 14H_2O \rightarrow 2Ca_3Al(OH)_6 \cdot 3SO_4 \cdot 26H_2O}$$

Both gypsum and ettringite cause expansion, create internal stresses, and lead to reduced structural capacity. The disturbance on the concrete microstructure by ettringite is similar to the occurrence of delayed ettringite formation, which is known to cause distress and cracking problems of concrete [8, 20].

In summary, the above mechanism of biogenic sulphuric acid attack is described by a four-stage process: first is the reduction of sulphate to sulphide by SRB; second is the disposition of hydrogen sulphide at the biofilm; third is the oxidation of sulphide to sulphuric acid by SOB; and fourth is the chemical attack of concrete by sulphuric acid [3]. Figure 1 illustrates the four stages. As the sulphuric acid at the concrete surface is concentrated, its actions can bring about serious deterioration of concrete, with the cementitious paste matrix being corroded and leaving behind the less acid-soluble aggregates imbedded in a soft, putty-like mass, as exemplified in Fig. 2.

In Hong Kong, a field survey had been conducted on a number of sewage treatment works by the Drainage Services Department. The survey revealed severe deterioration of concrete in various parts of the sewerage construction such as

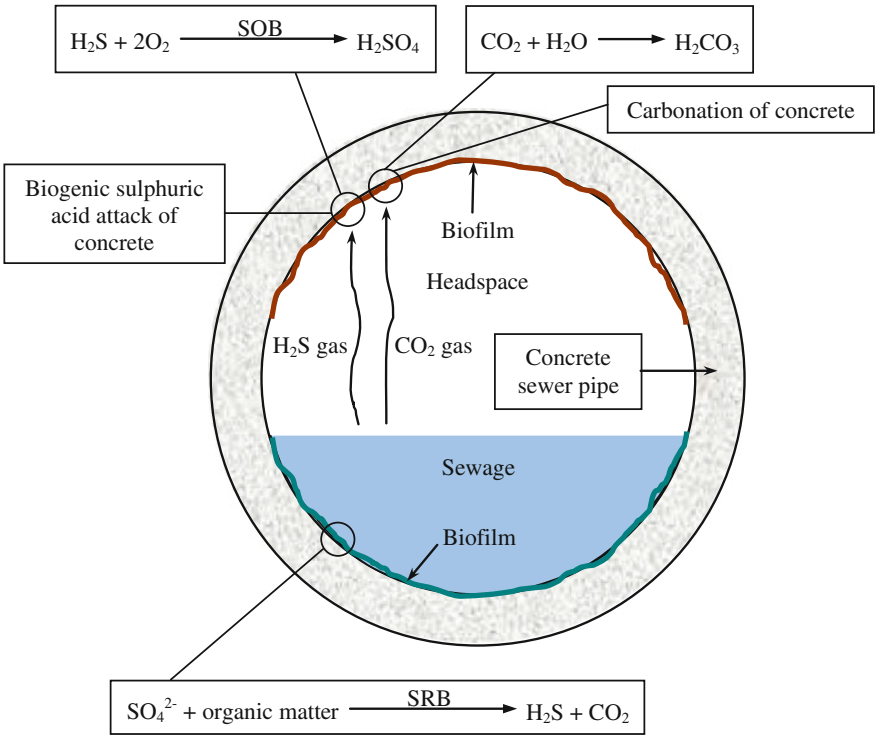


Fig. 1 Mechanism of biogenic sulphuric acid attack of concrete

Fig. 2 Deteriorated concrete by biogenic sulphuric acid attack

detention chamber, sludge thickening tank, and sedimentation tank. The phenomenon of paste matrix corrosion and disintegration showing exposed aggregate particles were readily observable [7]. The impact of biogenic sulphuric acid attack of concrete

sewerage systems in other countries and cities had been reviewed by Vincke et al. [30]. From worldwide practical experience, under favourable conditions of microbiological activities, the loss in concrete thickness can be up to a few millimetres per year [29, 30]. The deterioration of sewerage concrete necessitates extensive repairs and replacements. Therefore, the enhancement of concrete durability in sewerage applications can substantially reduce social costs and safeguard engineering assets.

3 Test Methods for Concrete Durability

The durability of concrete in sewerage environment may be evaluated by field tests and laboratory tests. The former includes field observation of sewerage construction and experimentation of concrete specimens after being mounted onto sewerage works in service for predefined period (Fig. 3) [33]. The latter includes chemical test and microbiological test [19]. The chemical test is performed by immersion of concrete specimens in sulphuric acid solution, whose concentration can be controlled to attain specific values of pH. The test can be carried out in two test conditions, namely close-to-reality condition where the pH value and temperature mimic the actual sewerage environment, and accelerated condition where the pH value is lower and the temperature is higher than reality. The microbiological test is performed by exposing the concrete specimens to an H_2S atmosphere to allow sorption of H_2S gas in the surface layer of concrete, and then immersing the specimens in a simulation suspension containing bacteria, sulphur and nutrients with incubation of SOB to produce biofilm and generate sulphuric acid. Subsequent to the chemical test or the microbiological test, the deterioration of concrete specimens may be visually assessed, and detailed inspection of the concrete microstructure with the use of scanning electronic microscope may be made [4]. The degree of concrete corrosion may be determined on the basis of the reductions of thickness and mass of the specimens.

Fig. 3 Concrete specimens mounted onto a sewer

4 Application of Protective Coatings

Protective coating provides a physical barrier to the hardened concrete against sulphuric acid attack. The effectiveness of protective coating is dependent on the material used. In practice, polymer modified cementitious coating (PMCC) and polymer resin coating are commonly adopted. PMCC is applied in the form of mortar coating, whose sulphuric acid resistant properties is improved by the addition of polymer to modify the microstructure of cement mortar [22]. A large variety of polymer resin coatings is available. They include epoxy resin coating, polyurea coating, polyurethane coating, and unsaturated polyester resin coating. De Muynck et al. [6] conducted chemical test and microbiological test for concrete cylinder specimens with measurement of the change in surface roughness, change in thickness, and mass loss of the specimens. It was found that the PMCC coating was not effective to protect the concrete from deterioration. On the other hand, epoxy resin coating and polyurea coating appeared to be effective in protecting the concrete.

Other types of coating may be applied. For example, silica sol coating which is composed of tetraethyl orthosilicate and ethanol, and hybrid coating which is produced by copolymerization of vinylacetate zirconium oxocluster and vinyltrimethoxy silane. According to Girardi [9], both silica sol coating and hybrid coating can reduce the mass loss of concrete specimens subjected to chemical test.

In spite of the favourable test results exhibited by some types of coatings, the anti-wash out property and the long-term adhesion of the coatings to concrete have not been researched. These are important performance attributes of protective coatings. Further investigation of these properties is necessary to ensure the continual functioning of protective coatings.

5 Mix Design to Improve Sulphuric Acid Resistance

The improvement of concrete durability against sulphuric acid attack via mix design has been extensively researched. Mix design here encompasses the selection of cementitious materials, pozzolanas, aggregates, fillers, admixtures, as well as their proportioning. These include the choice of type of cement, incorporation of fly ash (FA), silica fume (SF), metakaolin and blast-furnace slag, use of limestone aggregates, use of quartz powder filler and limestone filler, and addition of various polymeric additives and chemical admixtures.

Apart from ordinary Portland cement (OPC), other types of cement such as high-alumina cement, sulphate-resisting cement, and blended cement may be used in concrete production. De Belie et al. [5] tested different commercially manufactured sewers and reported that those produced from sulphate-resisting cement exhibited

higher durability. The influence of pozzolanas is dependent on the type used. Though chemical tests of mortar samples under pH values of 0.5, 1.0, 2.0 and 3.0, Jeon et al. [13] suggested that replacing 60 % of cement by ground granulated blast-furnace slag resulted in good sulphuric acid resistance, followed by replacing 20 % of cement by FA, followed by replacing 10 % of cement by SF, with the OPC mortar being least resistant to sulphuric acid. For FA and SF, the experiments performed by different researchers yielded widely different results. In line with Jeon et al. [13], Rahmani and Ramazanianpour [23] found that replacing 8 % of cement by SF yielded marginal improvement of concrete durability. Soroushian et al. [25] reported that the sulphuric acid resistance of concrete was enhanced by FA and SF additions. On the contrary, Vincke et al. [31] discovered through chemical and microbiological tests that adding SF had adverse effect on the concrete durability. Cizer et al. [4] reported adverse effect on sulphuric acid resistance by addition of FA. Hewayde et al. [11] found that the effect of SF addition was minor, whereas the incorporation of metakaolin in concrete was effective in improving the sulphuric acid resistance.

Positive results have been reported regarding the use of limestone material in concrete production. Through chemical tests with sulphuric acid solutions under pH values of 1.0, 2.0 and 3.0, Bassuoni et al. [2] discovered that the use of limestone coarse and fine aggregates improved the sulphuric acid resistance of concrete under highly acidic condition (pH value of 1.0), and the use of limestone filler improved the sulphuric acid resistance of concrete at pH value of 2.0. Cizer et al. [4] also recognized that the use of limestone filler was effective. However, conflicting results had been obtained regarding the use of quartz powder filler. Rahmani and Ramazanianpour [23] found that the concrete incorporating ultra-fine quartz powder filler was more resistant to sulphuric acid and they attributed this to the denser particle packing. In contrast, Cizer et al. [4] reported adverse effect of using quartz powder filler and they related such finding to the pore densification effect which created less space to release the internal stresses.

Various polymeric additives could be added to the concrete for modifying its properties to enhance durability. Monteny et al. [18] and Vincke et al. [31] investigated different types of polymer-modified concrete through chemical and microbiological tests, and reported that the addition of styrene-acrylic ester polymer could slightly increase the biogenic sulphuric acid resistance of concrete, the addition of acrylic polymer could slightly decrease the biogenic sulphuric acid resistance, while the addition of styrene butadiene polymer led to insignificant effect or slight decrease of resistance, and the result of adding vinylcopolymer was not significant. Soroushian et al. [25] tested concrete with addition of vinyl acetate-ethylene copolymer and found improvement of sulphuric acid resistance.

Furthermore, the addition of chemical admixtures was explored. Hewayde et al. [11] discovered that adding calcite (ammonium stearate solution) and adding xypex (a proprietary waterproofing admixture for concrete) were both effective means to

improve the resistance against sulphuric acid attack. It should be noted that both calcite and xypex would undergo crystalline reactions and block pores in concrete to effectuate the waterproofing performance. Besides, De Muynck et al. [6] investigated the addition of hydrous silicate admixture but the result was not effective.

Considering the above literature findings, it can be seen that the experimental results regarding addition of pozzolanas were diverse, and the results with limestone aggregates and limestone filler were appealing. Further research is recommended to confirm the effects of adding various pozzolanas, and to formulate mix design strategies to optimize the use of limestone aggregates and limestone filler. Besides, the use of pore blocking or other specific chemical admixtures to enhance sulphuric acid resistance could be an emerging technology.

6 Use of Corrosion Inhibitors

Corrosion inhibitors, in the context of this chapter, refer to the deactivating chemicals that suppress the bacteria activities in connection with sewerage concrete corrosion. Corrosion inhibitors can be in the form of biocides or reagents added directly to the concrete mixture, coated onto fibres mixed with the fresh concrete, or sprayed onto the sewer pipe. Hewayde et al. [11] tested concrete specimens produced with the addition of organic corrosion inhibitor which was an aqueous mixture of amines and esters. It was found that the deterioration of concrete was substantially ameliorated. On the other hand, De Muynck et al. [6] studied the effectiveness of adding anti-microbial fibres treated with biocides and adding anti-microbial silver/copper zeolites into the concrete mixes. The experimental results showed that both the anti-microbial fibres and the anti-microbial zeolites could suppress bacterial activity but could not protect the concrete against deterioration. The addition of metal inhibitors such as selenium and nickel in concrete was recognized to be able to suppress SOB activity, by binding to the bacteria cells and inhibiting enzymes involved in oxidation of sulphur [30].

Spraying of corrosion inhibitors onto sewer pipes can be performed through the crown spray process, whereby a float mounted with a spray head is pulled along the sewer at a controlled rate to spray the crown at a predetermined application rate [27]. Alkaline chemicals can be sprayed to neutralize the acid generated on the concrete [22]. Application of biocides such as phenol derivatives, quaternary ammonium salts, and monochloramine by the crown spray process had been researched. The use of magnesium hydroxide in 50 % concentration was found to offer effective and prolonged protection to concrete [30]. Notwithstanding this, it should be noted that the corrosion inhibitor sprayed onto the sewer would be subjected to wash-out by the sewage flow. Re-spraying can be applied in suitable time intervals to restore the effectiveness of protection over a period, and this adds to the maintenance cost of the sewerage system. Further research on the anti-wash out property of sprayed corrosion inhibitors is needed.

7 Conclusions

The sewerage environment poses an immense challenge to the durability of concrete, due mainly to the generation of concentrated sulphuric acid at the biofilm on concrete surface by microbial activities. In this chapter, the mechanism of biogenic sulphuric acid attack of concrete has been elucidated. Test methods to assess the resistance of concrete against such attack have been highlighted. Various possible ways of improving the durability of concrete of sewerage infrastructure, including the use of protective coatings, better concrete mix design to improve the biogenic sulphuric acid resistance of concrete, and use of corrosion inhibitors have been discussed. A literature review of experimental investigations of the effectiveness of different measures has been conducted. The authors recommend further research on developing an effective and robust method for durability enhancement of sewerage concrete, so as to reduce the life-cycle cost and protect the engineering asset of the sewerage system. Last but not least, despite taking measures to achieve more durable concrete, reasonable inspection and maintenance over the service life are deemed required, and the cost associated with such should be accounted for in the life-cycle cost analysis of the sewerage construction.

References

1. Alexander M, Mackechnie J (2003) Concrete mixes for durable marine structures. J South Afr Inst Civil Eng 45(2):20–25
2. Bassuoni MT, Nehdi M, Amin M (2007) Self-compacting concrete: using limestone to resist sulfuric acid. Proc Inst Civil Eng Constr Mater 160(3):113–123
3. Beeldens A, Van Gemert D (2001) Biogenic sulphuric acid attack of concrete sewer pipes: a prediction of the corrosion rate. In: Malhotra VM (ed) Proceedings, fifth international conference on recent advances in concrete technology, ACI SP-200. American Concrete Institute, pp 595–606
4. Cizer O, Elsen J, Feys D, Heirman G, Vandewalle L, Van Gemert D, De Schutter G, Desmet B, Vantomme J (2011) Microstructural changes in self-compacting concrete by sulphuric acid attack. In: Proceedings, 13th international congress on the chemistry of cement, Madrid, Spain, pp 436–442
5. De Belie N, Monteny J, Beeldens A, Vincke E, Van Gemert D, Verstraete W (2004) Experimental research and prediction of the effect of chemical and biogenic sulfuric acid on different types of commercially produced concrete sewer pipes. Cem Concr Res 34 (12):2223–2236
6. De Muynck W, De Belie N, Verstraete W (2009) Effectiveness of admixtures, surface treatments and antimicrobial compounds against biogenic sulfuric acid corrosion of concrete. Cem Concr Compos 31(3):163–170
7. Drainage Services Department (2008) Concrete design for sewerage structures. Report No. RD 1055, Sewerage Projects Division, Drainage Services Department, Hong Kong, p 26
8. Famy C, Taylor HFW (2001) Ettringite in hydration of Portland cement concrete and its occurrence in mature concretes. ACI Mater J 98(4):350–356

9. Girardi F (2009) Studies on concrete degradation in aggressive environment and development of protective system. Department of Materials Engineering and Industrial Technologies, University of Trento, p 111
10. Grace WR (1988) Durable concrete for marine structures—guidelines for designers. Hong Kong Eng 16(3):17–22
11. Hewayde E, Nehdi ML, Allouche E, Nakhla G (2007) Using concrete admixtures for sulphuric acid resistance. Proc Inst Civil Eng Constr Mater 160(1):25–35
12. Jensen HS (2009) Hydrogen sulfide induced concrete corrosion of sewer networks. PhD thesis, Aalborg University, Denmark, p 67
13. Jeon JK, Moon HY, Ann KY, Kim HS, Kim YB (2006) Effect of ground granulated blast furnace slag, pulverized fuel ash, silica fume on sulfuric acid corrosion resistance of cement matrix. Int J Concr Struct Mater 18(2):97–102
14. Kalhøj M (2009) Hydrogen sulfide oxidation and sewer corrosion. Department of Biotechnology, Chemistry and Environmental Engineering, Aalborg University, Denmark, p 40
15. Kwan AKH, Wong HHC (2005) Durability of reinforced concrete structures: theory and practice. In: Proceedings, Hong Kong government standing committee on concrete technology annual concrete seminar, Hong Kong, pp 1–20
16. Marquez JF, Sanchez-Silva M, Husserl J (2013) Review of reinforced concrete biodeterioration mechanisms. In: Proceedings, 8th international conference on fracture mechanics of concrete and concrete structures, Toledo, Spain, p 9
17. Metcalf & Eddy (2003) Wastewater engineering: treatment and reuse, 4th edn. McGraw-Hill, Boston, pp 18–19
18. Monteny J, De Belie N, Vincke E, Verstraete W, Taerwe L (2001) Chemical and microbiological tests to simulate sulfuric acid corrosion of polymer-modified concrete. Cem Concr Res 31(9):1359–1365
19. Monteny J, Vincke E, Beeldens A, De Belie N, Taerwe L, Van Gemert D, Verstraete W (2000) Chemical, microbiological, and in situ test methods for biogenic sulfuric acid corrosion of concrete. Cem Concr Res 30(4):623–634
20. Neville AM (2011) Properties of concrete, 5th edn. Pearson Education, Harlow, p 846
21. O'Connell M, McNally C, Richardson MG (2010) Biochemical attack on concrete in wastewater applications: a state of the art review. Cement Concr Compos 32(7):479–485
22. Parande AK, Ramsamy PL, Ethirajan S, Rao CRK, Palanisamy N (2006) Deterioration of reinforced concrete in sewer environments. Proc Inst Civil Eng Munic Eng 159(1):11–20
23. Rahmani H, Ramazanianpour AA (2008) Effect of binary cement replacement materials on sulfuric acid resistance of dense concretes. Mag Concr Res 60(2):145–155
24. Skalny J, Marchand J and Odler I (2002) Sulfate attack on concrete. Spon Press, London, p 217
25. Soroushian P, Nassar RUD, Chowdhury H, Ghebrab T (2010) Testing concrete durability in sewer environment. Proc Inst Civil Eng Constr Mater 163(1):35–44
26. Suprenant BA (1991) Designing concrete for exposure to seawater. Conc Constr 36(12)
27. Sydney R, Esfandi E, Surapaneni S (1996) Control of concrete sewer corrosion via the crown spray process. Water Environ Res 68(3):338–347
28. Trejo D, De Figueiredo P, Sanchez M, Gonzalez C, Wei SP, Li L (2008) Analysis and assessment of microbial biofilm-mediated concrete deterioration. Texas Transportation Institute, Texas, USA, p 26
29. Van Mechelen T, Polder R (1997) Biogenic sulphuric acid attack on concrete in sewer environments. In: Proceedings of the international conference on the implications of ground chemistry and microbiology for construction, Bristol, UK, pp 511–524
30. Vincke E, Monteny J, Beeldens A, De Belie N, Taerwe L, Van Gemert D, Verstraete W (2000) Recent developments in research on biogenic sulfuric acid attack of concrete. In: Lens PNL, Hulshoff Pol L (eds) Environmental technologies to treat sulfur pollution: principles and engineering. IWA Publishing, London, pp 515–541

31. Vincke E, Wanseele EV, Monteny J, Beeldens A, De Belie N, Taerwe L, Van Gemert D, Verstraete W (2002) Influence of polymer addition on biogenic sulfuric acid attack of concrete. Int Biodeterior Biodegradation 49(4):283–292
32. Wei SP, Sanchez M, Trejo D, Gillis C (2010) Microbial mediated deterioration of reinforced concrete structures. Int Biodeterior Biodegradation 64(8):748–754
33. Wells T, Melchers RE, Bond P (2009) Factors involved in the long term corrosion of concrete sewers. In: Australasian corrosion association proceedings of corrosion and prevention, Coffs Harbour, Australia, p 11

Successful Reduction of Non-revenue Water (NRW)

Sheng JIN and Jinghui TANG

Abstract Sino French Water, affiliated company of Suez Environnement, invests and establishes joint ventures (JVs) with local water companies in China to improve their managerial and operational efficiencies. There are more than 26 Sino French JVs spread all over China. The success story of NRW reduction from 35 to 5 % in Tanzhou took 18 years. This success is largely due to Suez Environnement global approach and expertise on NRW management. The successful actions performed in Tanzhou can be divided into five phases and include both managerial and technical changes in the Company. (1) One of the first actions taken in 1995 was the replacement of customer meters that were previously randomly chosen. The replacement of these meters by accredited meter suppliers and standardizing supplier selection lead to drop in NRW of 22 % in two years. (2) From the success of the first two years, a set of management procedures and standardizations were put in place in 1997. Additional income from accurate metering allows investing in a network rehabilitation, which brought the NRW further down to 13 % by 2002. (3) Between 2002 and 2005 the NRW actually went up. Investigations in the production meter and the establishment of a leak detection and valve maintenance team finally helped stabilize the NRW situation at around 12.5 % by 2005. (4) Applying Aqua Circle tool since 2007 allowed Tanzhou to further understand NRW and how other indicators, besides percentage, were critical. (5) Between 2007 and 2013, a combination of advance technologies (DMAs, PMAs, advance acoustic loggers, Helium gas) and management systems (GIS, hydraulic model) has driven Tanzhou to an extremely low, yet sustainable NRW of 4 %. Tanzhou experience is a valuable

S. JIN (✉)
Network & Customer Services Department, Sino French Water, Shanghai, P.R. China
e-mail: jinsheng@sinofrench.com

J. TANG
Zhongshan Tanzhou Water Supply Co., Ltd, Sino French Water, Shanghai, P.R. China
e-mail: tangjh@sinofrench.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_90

example of NRW reduction and asset management for Sino French Water and Suez Environnement. Reducing and maintaining a low NRW is a constant challenge and never-ending activity.

Keywords NRW · China · Sino french water · Suez environnement

1 Introduction

Sino French Water is a company created in 1992 between Suez Environnement of France and NWS Holdings Limited of Hong Kong. Sino French Water invests and establishes joint-ventures (JVs) with local water authorities and companies in China to improve the management and operation of water, waste water, industrial water services and sludge treatment. Today there are more than 25 JVs spread all over China and each one with unique characteristics. With such diverse conditions, systematic methodologies are put in place but applied in a pragmatic manner depending on the condition and maturity of the local company. This chapter will explain the various actions taken by Sino French Water in five different phases over 18 years to reduce the non-revenue water (NRW) from 35 to 5 % at its JV in Tanzhou, China (Fig. 1).

Zhongshan Tanzhou Water Supply Company Limited (Tanzhou Water) is the first joint-venture of Sino French Water in mainland China and it is also the first Chinese-Western joint-venture in the water business in China.

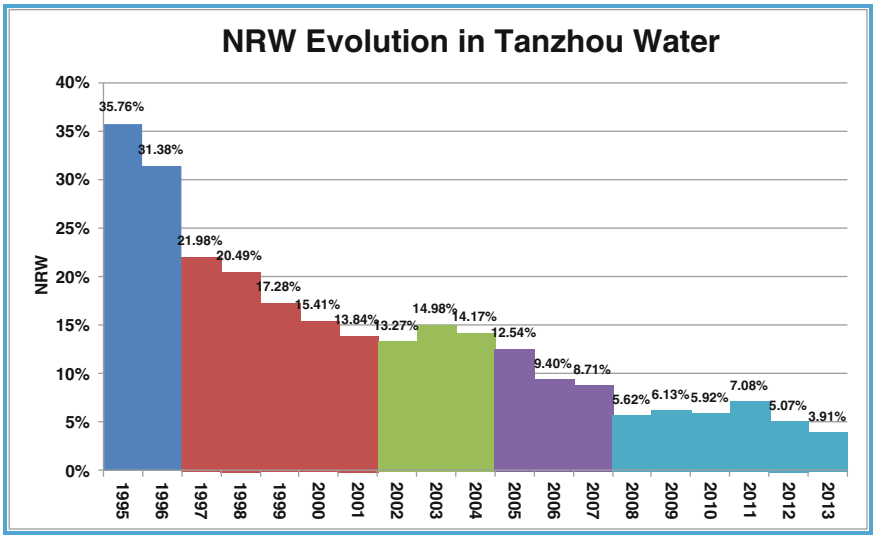


Fig. 1 NRW evolution of Tanzhou water

Fig. 2 Map of Hong Kong region, with the location of Tanzhou

Tanzhou is located in Guangdong province approximately 45 min from the border with Macao, a Portuguese colony for more than 400 years and handed over back to China in 1999. It is important to note that the supply of water in Macao is managed since 1985 by the Macao Water Supply Company (Macao Water) another subsidiary of Suez Environnement and NWS Holding. Its proximity to Macao has allowed the people of Tanzhou Water to closely witness the improvements made in Macao Water with the introduction of Suez Environnement methods and techniques and to quickly adopt the effective experiences and thus benefit significantly from being part of an international group (Fig. 2).

2 Phase 1 [1995–1997]—Quick Wins

After the establishment of the joint-venture between Sino French Water and the local government, one of the key objectives for Tanzhou Water was to reduce NRW and to optimize asset efficiency and revenue. Four key processes were put in place: rationalized meter replacement, inspection and repairs, leak report awarding system, and leak detection using check meters.

2.1 Rationalized Meter Replacement

Customer meters prior to 1995 were selected arbitrarily from various suppliers. The meters were not tested, so the metrology was questionable and the performance not guaranteed. In addition, no meter replacement had taken place between 1981 and 1995.

Thus starting from the end of 1995, Tanzhou Water started a company-wide program to replace all customer meters. In order to better manage their meter population and to guarantee their performance, Tanzhou Water only bought meters from reliable suppliers, which they limited to one or two for the ease of replacement and repair. Guided by the experience from Macao Water, the subsidiary of Suez Environnement, Tanzhou Water set up their own meter test centre in 1996 using the comparison method to verify meter performances.

To avoid under-metering losses due to ageing, a meter replacement policy was also introduced to replace all meters every 3 years.

2.2 Inspection and Repairs

The second action was the setup of a dedicated inspection team responsible for surveying the network that had dual benefit. First of all, at that time, leaks ran for a long time before anyone reported it and secondly, there was a serious fraud problem in the rural areas. By creating a dedicated inspection team which walked along the pipelines, Tanzhou Water greatly reduced leakage reporting time as well as the number of fraud cases.

To further reduce leak reporting time, the leak reporting awarding policy has been setup in 1995. This action not only motivated company staff to report leaks, but helped them develop a habit and a culture of reporting leaks. With the establishment of the inspection team and frequent network surveillance, the number of fraud cases naturally decreased.

In the area of repairs, Tanzhou water did not need to start from scratch. There was already a repair team. The repair techniques of the team were to be improved, as well as their transportation means (see Fig. 3). The average repair time was too long, between 4 and 8 h. By reinforcing the size of the repair team, equipping them with the proper transportation vehicles, tools and equipment, the repair time was greatly reduced to an average of 2–4 h.

2.3 Leak Detection by Check Meter

Beside meter testing using a standard meter, Tanzhou Water also learned from Macao Water about organizing leak detection by areas using check meters. Since there was no dedicated leak detection team, this method was useful in optimizing their efficiency. Before 1995 in Tanzhou, leak detection was done by using listening sticks and stethoscopes and there were no systematic leak detection method. Macao Water provided technical support to train Tanzhou Water team 2 to 3 times per year, 1 week in duration each time. Together they identified 11 branched networks with single inlet where they installed check meters. By reading the meter and by taking manual night flow measurements, they were be able to determine if there

Fig. 3 Method of transportation of the repair team before and after

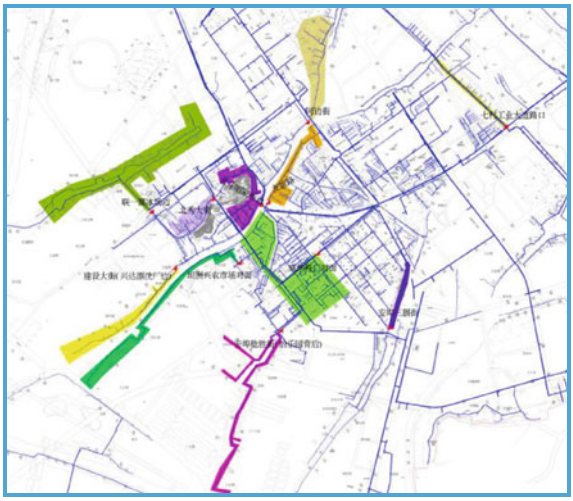


Fig. 4 The 11 check meter zones

were any leaks in the network, then carry out leak detection. This method helped Tanzhou Water leak detection team to narrow down the search area for leaks, increasing their efficiency as well as the confidence in their skills (Fig. 4).

2.4 Management

One crucial point to the success of these actions was the combination of experienced managers from Suez Environnement and motivated high potential local managers from Tanzhou Water. Suez Environnement managers had the experience of implementing these actions elsewhere before so they were able to anticipate the problems that might arise while deploying the actions and ensure the smooth operation and proper execution of these projects. Local managers were trained to develop further and maintain the implementation of the actions.

3 Phase 2 [1997–2002]—Documentation and Network Investment

3.1 Process Documentation

After the first 2 years of dramatic changes and hard work, the NRW of Tanzhou dropped from 35.76 % in 1995 to 21.98 % in 1997. These results proved that the actions put in place were effective. Thus the next step was to document and standardize the processes, so that they would become long-term practices.

Some of the work flow procedures established included leak detection, inspection, maintenance, network design, construction, and acceptance. Only writing these procedures was not enough to give the staff ownership of what they are doing. So in 1999 Tanzhou Water extended these procedures and wrote down the responsibility for each position. This allowed each staff to know exactly what their responsibilities were and what was expected of them. These documentations greatly improved the management allowing for smoother work communication and increasing efficiency.

These procedures marked the transition of management from foreign to local managers. Having worked alongside foreign managers and having a set of work flow documentation, by 2000 all departmental managers were local mainland Chinese.

3.2 Network Rehabilitation

With the additional income generated from accurate metering, investment in network rehabilitation was possible. At that time network rehabilitation was a passive

action, pipes would only be replaced if there were so many leaks that it could not be repaired anymore. It was a major problem because the pipe materials that were used were poor: they were mainly concrete, galvanized iron (G.I.), and cast iron pipes with rigid joints. In the 1980s, until 1998, there were many leaks and bursts due to pipe corrosion and blocked pipes.

In the beginning of 1999, a network rehabilitation and replacement plan for the entire network was drafted and there has been a yearly network rehabilitation plan and budget for rehabilitation every year since 2000.

With the problems that it had with the pipe materials, Tanzhou Water started to look for alternative pipe materials. In 2000, they replaced cast iron pipes with ductile iron pipes and in 2002 they replaced G.I. pipes by steel pipes with plastic lining. The combination of network rehabilitation and the use of alternate pipe material reduced the frequency of pipe bursts and improved the condition of water supply. By the end of 2002, the NRW in Tanzhou was 13.27 %.

4 Phase 3 [2002–2005]—Extending the Program to Other Areas and Keeping up with Technology

4.1 Maintenance of Production Meters

In 2003 the NRW increased to 14.98 % and one of the first things that was checked was the production flow meter. The production flow meter was an insertion type single-beam ultrasonic flow meter made in China and large fluctuations were started to be observed in its measurement.

As a result, Tanzhou Water decided to replace the ultrasonic flow meter with an electromagnetic flow meter from Shanghai Krohne. In addition to the replacement of the meter, Tanzhou also set up a flow meter maintenance policy.

By replacing the ultrasonic flow meter, the NRW dropped from 14.98 % in 2003 to 14.17 % in 2004. From this experience, the team of Tanzhou Water saw the significance of the accuracy of the flow meter and production volume. So to minimize month to month variation and meter reading lag, they have set up a policy to manually read the raw and production flow meters on a particular day and a particular time each month and use it as the data for NRW calculation.

4.2 Establishment of Valve Team and Leak Detection by Acoustic Equipment

As Tanzhou's network started to age, the valves started to have problems. Through repair feedback, it was learnt that many of the main valves in the network could not be completely closed leading to additional water loss and increasing repair time. As

a result, an additional team, the valve management team, was set up in the distribution department. The valve management team would be responsible for checking, repairing, and the replacement of valves on mains.

In addition to valve management, leak detection also needed to improve to further reduce NRW and to keep up with the ageing network; thus Tanzhou Water decided to create an active leak detection team and invited the Macao Water leak detection team to conduct training for them. From the training, Tanzhou Water realized that the equipment they were using, listening sticks and stethoscopes were outdated and needed to purchase new and more advanced equipment which included noise loggers and correlators in order to increase their efficiency and competence to pinpoint leaks.

5 Phase 4 [2005–2007]—Deepening Understanding of NRW Calculation and NRW Indicators

5.1 Bulk Water Supply

Starting in May 2005, Tanzhou started bulk water supply to a nearby village, Sanxiang. This is Tanzhou Water's largest customer. Without significantly extending the network, Tanzhou increased its water sales by approximately 10 %.

Even though Tanzhou Water did not need to deploy additional efforts to improve NRW, the NRW ratio improved from 12.54 to 8.71 %, and this gave the staff of Tanzhou Water insight to the various components and factors that can affect NRW, how other indicators, like volume loss per kilometre of network, besides NRW ratio, are also important to assess the effectiveness of the NRW program. A further study using Suez Enviornnement Aqua Circle tool to analyze NRW components and network performance was carried out with the assistance of Suez Enviornnement expert so as to produce a targeted action plan for NRW management (Fig. 5).

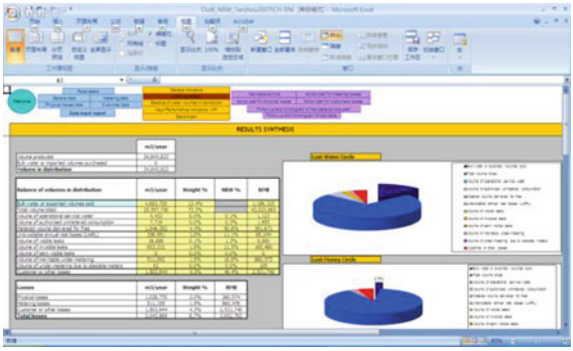


Fig. 5 Aqua circle tool of suez environnement

6 Phase 5 [2007–2012]—Reaching World-Class Performance

This last phase is the most remarkable and significant. It demonstrates how Tanzhou Water is a mature water company with practices comparable with water companies in Western, developed countries.

6.1 Establishment of a Company-Wide NRW Team

Prior to 2007, actions contributing to the reduction of NRW were implemented independently by each department and there was little inter-departmental communication. From the experience of production meter influence and upon the recommendation of the Technical Department of Sino French Head Quarters (SFHQ), Tanzhou Water established a NRW team composed of the following: a team leader, a representative from the Water Supply Department (WSD), a representative from the Customer Services Department, and a representative from the Distribution Department.

The team leader is responsible for coordinating and overseeing the NRW program, drafting action plan, communicating with the company regarding the action plan and negotiating for budgets. The representative from the Water Supply Department (WSD) is responsible for organizing all activities in WSD, for example, monitoring and controlling pressure, checking the water level of reservoirs, and validating production data. The Customer Service Department (CSD) representative is responsible for activities in CSD such as meter reading, metering, setting up check meter systems, and analyzing customer data. Finally, the Distribution Department representative is responsible for activities in distribution such as consolidating and analyzing leakage data, setting up network zones, night flow activities, and repairs.

6.2 Metering Policy

The metering policy is an initiative from SFHQ. Its aim is to standardize meter management within all Sino French Water JVs and to ensure that all the meters used are approved and of good quality. The policy controls the entire life cycle a water meter from meter selection, purchasing, testing, acceptance, in-service meter management, to disposal; however, at the same time allowing a certain degree of flexibility for the individual JVs.

This policy serves as a general structure for Tanzhou Water's metering management. Previous documentations and ISO procedures regarding meters are a complement to this framework. Having this policy further enhances meter management in Tanzhou to minimize commercial and metering losses.

6.3 Geographic Information System (GIS)

In 2007, Tanzhou invested in the GIS system, the software and the pipeline survey of which 280 km were outsourced and 150 km were surveyed in-house by Tanzhou Water GIS team.

During the implementation of the GIS, Tanzhou Water worked with a subcontractor while developing an internal professional team dedicated to pipeline survey, composed of 3 to 4 people. This pipeline survey team is responsible for pipe detection and localization, coordinates measurement, network data verification, and uploading data to GIS database.

With the implementation of the GIS system, Tanzhou Water, created a set of documentation to control and maintain the system. The documentation includes: GIS management, as-built drawings acceptance standards, operation manual of pipeline survey, and forms for recording pipe failures.

It is not difficult to set up a GIS system, many water companies have one. However, the key is in the application of the system and how the water company can take the most out of it to their advantage. The following is a list of the most common actions used: inquiry and statistical calculations, network input and edit, base map management, valves shut-off, pipe failure management, project information management, and DMA (District Metering Area) management. This GIS system greatly helped Tanzhou Water's NRW program: pipe failure management provided accurate data to support the need to replace and rehabilitate the network; the valve shut-off function indicated which valves to shut off on the field avoiding the time wasted from trying to close mal-functioning valves and reducing the time to search for valves on the field; night flow data from the loggers on the check meter helped the management and analysis in organizing leak detection activities. In the long term, by building the GIS and inputting all network information, the GIS became the basis of more effective network asset management and an integral part of sustaining a lower level of NRW.

Additionally, the GIS contains meter information which is connected to the customer database. Therefore when there is a pipe burst, a list of affected customers could be generated for the Customer Services Department to contact. In addition, Tanzhou Water is also planning to plot down customers with various types of complains to better understand and address their customers' needs.

6.4 Hydraulic Model

Prior to December 2007, Tanzhou Water didn't have its own hydraulic model and for the occasional master planning, Tanzhou Water received support from Macao Water.

As Tanzhou Water's network grew and the company matured, they could not always rely on the support of Macao Water. Therefore in December 2007, they invested in the hydraulic modelling software and corresponding training. Since the purchase of the software, Tanzhou Water first developed a planning model in 2008, then a more accurate static model in 2009, and then a dynamic model in 2010. Now, the hydraulic model is part of Tanzhou Water's standard operations. The area of application includes: network assessment, master planning, network design, network renovation, network sectorization, energy savings, water age analysis, pressure management and emergency planning, of which, sectorization and pressure management are particularly important to NRW control. Tanzhou Water first uses the hydraulic model to sectorize and create DMAs in the network to make sure that the zones are hydraulically sustainable afterwards; the DMAs are managed in the GIS.

6.5 Leak Detection Equipment and Strategy

In the first quarter of 2009, some of the noise loggers in Tanzhou Water started to breakdown and were sent back to the manufacturer for repairs. During this time, the number of invisible leaks found in Tanzhou dramatically decreased and the monthly NRW started to rise above 10 %. As a result Tanzhou looked into purchasing new and advance leak detection equipment. The new noise loggers had correlation function which could further help narrow down the area of suspected leak and the correlator could make a correlation in within 5 min, while the current equipment required 20 min.

So in June 2009, Tanzhou Water received 2 sets of new leak detection equipment for a one month trial. Working overtime to take maximum advantage of these equipments Tanzhou Water found 37 invisible leaks that month, bringing the monthly NRW down to 6.7 %. The trial of the equipment was a great success, the NRW percentage and the number of invisible leaks detected clearly supported the success (Fig. 6).

In addition to acoustics leak detection equipment, Tanzhou Water also looked into new leak detection technology: helium, which was a new technology for the water industry in China. This was a portable helium leak detection unit developed within the Suez Environnement group.

In the past 2 years, with network expansion, Tanzhou Water has laid over 200 km of PE pipes. The acoustic leak detection on non-metallic pipes was not very effective. So originally Tanzhou Water purchased the helium equipment for this

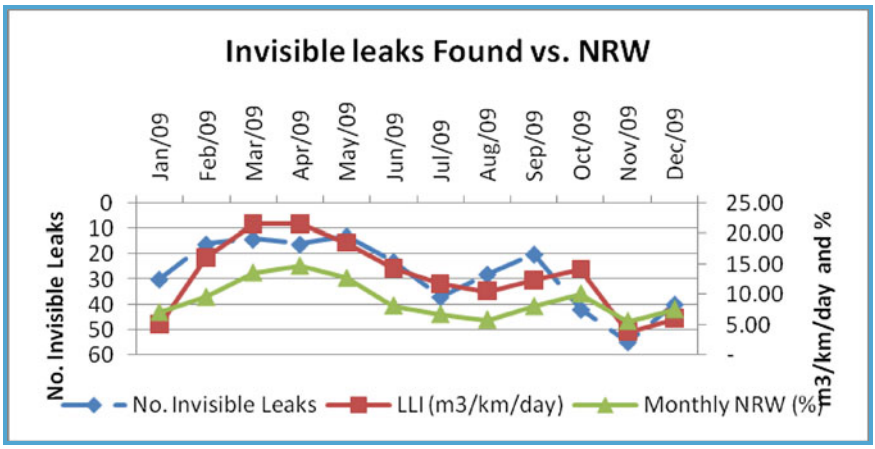


Fig. 6 Relationship between NRW and invisible leaks found

purpose. However, since these pipes are new, they do not have many leaks, so now they use it on metallic network as well, particularly in areas where acoustic equipment doesn't work very well, such as noisy areas, industrial zones, and areas where the network is complicated and correlation is difficult. The equipment is able to pinpoint the leak with high precision and reliability (Fig. 7).

With the addition of new equipment but needing to keep the same number of staff, Tanzhou Water needed to revise their leak detection plan to optimize their resources. Considering the network conditions, the leak detection team decided to use the most suitable leak detection equipment for each area. From the figure below, the solid polygonal pink and green areas are most suitable for noise loggers; the

Fig. 7 Helium detection system

Fig. 8 Leak detection strategy—equipment suitability

orange circular area is an industrial zone, which is suitable for helium gas; finally blue dotted area is suitable for geophone and listening sticks as there are not many valves on the pipe as exposure points (Fig. 8).

6.6 Zones and DMAs

With the recent growth in Tanzhou's network, they needed to restructure the DMAs. So with the construction of the hydraulic model, Tanzhou Water took this opportunity to purchase real-time flow and pressure loggers to set up some big sectors. Using 8 flow meters, they divided the entire Tanzhou network into 4 big zones. Within these 4 zones, there are about 30 DMAs (see Fig. 9). With these real-time data, night flow monitoring could be done more easily and efficiently. In some sectors, not all the network can be further split into smaller DMA at the moment due to pressure and water quality constraints. However, Tanzhou Water is working towards this gradually through network rehabilitation and replacement.

Meter information is collected in the GIS from the customer database. So Tanzhou Water uses the monthly meter reading data from CSD to manage their DMAs and to prioritize leak detection activities.

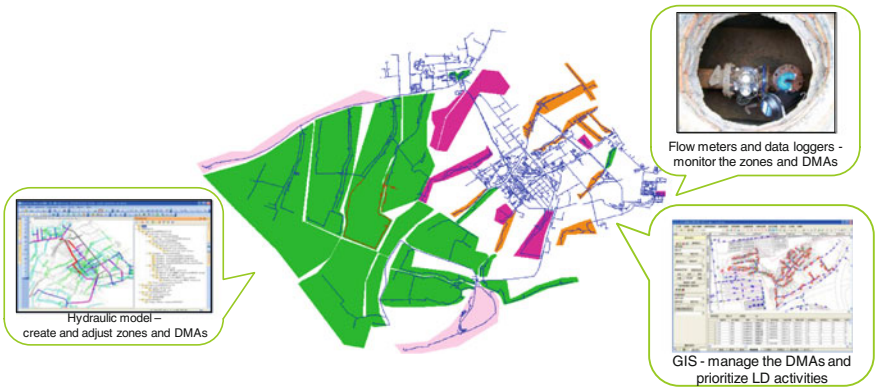


Fig. 9 DMAs in Tanzhou in 2010

7 Conclusion

From the experience in Tanzhou Water, it can be seen that technical improvement is an important part in reducing NRW; however it must be complemented at the right time with proper management to sustain and continue its improvement. Having a step-by-step approach is critical to its success. If the documents and policies were put in place at the very beginning of the establishment of the JV they might not have been as effective since the knowledge and trust in the new practices were not established among the local teams making it difficult for them to understand the purpose and the importance of procedures. They could even develop a sense of resistance against paperwork and the initiatives launched by the management. In other words, launching proper initiatives at the appropriate time according to the development and maturity of the local teams and management is of utmost importance.

Over 18 years, there had been five stages during which Tanzhou Water progressively enhanced their leak detection. This was done step-by-step, according to the maturity of the leak detection team. If noise loggers or even helium were introduced in the very beginning, they might not have developed the best practice of separating the network into DMAs and prioritizing leak detection activities.

The success of Tanzhou Water relied heavily on the motivation and determination of local staff. However, being in an international group such as Suez Environnement allowed them to accelerate this success due to the support from mother and sister companies and exposures to best practice and new technologies.

Now Tanzhou Water is not only a reference for other JVs within Sino French and subsidiaries in the Suez Environnement group, it is often visited by other water companies in China.

Identifying Key Performance Indicators for Engineering Facilities in Commercial Buildings—A Focus Group Study in Hong Kong

C.S. Man and Joseph H.K. Lai

Abstract Hong Kong is a vibrant city with numerous commercial buildings that are large in scale and good in quality. Keeping or enhancing the value of these buildings in the long run relies on the satisfactory performance of their engineering facilities. To ensure the effective management of the operation and maintenance (O&M) of these facilities, it is necessary to measure their input resources and outcome performance. Since a holistic scheme that can evaluate the performance of engineering facilities in commercial buildings in Hong Kong was not available, a research study has been undertaken in an attempt to develop such a scheme. Forming part of the study, a focus group meeting that aimed at identifying key performance indicators (KPIs) for representing the performance of the engineering facilities was convened with the participation of highly experienced and professional O&M experts. This paper reports on the arrangement of this focus group study, the design of its data collection tool, the process of the discussion among the experts, and the ways in which the KPIs were shortlisted. Future works needed for completing the development of the scheme are also described.

1 Introduction

Hong Kong is packed with a high density of buildings, accommodating over 7 million people in a limited land area of 1,104 km^2 [1, 2]. It is also a glamorous city attracting over 48 million visitors a year [3]. With such a large number of residents and visitors, it is not difficult to imagine how important the building sector is, especially the commercial buildings where a variety of business and leisure

C.S. Man (✉) · J.H.K. Lai
Department of Building Services Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
e-mail: m.c.sing@connect.polyu.hk

J.H.K. Lai
e-mail: joseph.lai@polyu.edu.hk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_91

activities take place. Commercial buildings in Hong Kong are well-known for their high sale and rental values. Keeping or enhancing the value of these buildings in the long run relies on satisfactory operation and maintenance of their engineering facilities [4], such as air-conditioning, electrical and fire services installations. Inadequate performance of the facilities may adversely affect the commercial operations of the buildings and even lead to enormous financial losses, impacts on the environment, or health and safety threats to the building end users.

Proper operation, maintenance and management of the engineering facilities are therefore vital. To achieve these, substantial amounts of input resources are needed. Since budgeted resources for building operation and maintenance are typically limited [5], priority setting is often needed for making decision on the allocation of resources [6, 7]. Hence, it is necessary to evaluate the performance of engineering facilities and assess the effectiveness of the input resources. Performance evaluation on engineering facilities enables us to monitor the output quality of works and identify any need for improvement before setting these priorities.

There were studies or assessment schemes pinpointing at some performance aspects of engineering facilities in existing commercial buildings, e.g. energy or environmental performance. However, a holistic scheme that measures the performance of engineering facilities in a wide range of operation and maintenance facets is yet to be seen in Hong Kong. In view of this, a study was commenced, which aimed to develop a performance assessment scheme for engineering facilities in commercial buildings in Hong Kong.

The initial phase of the study was completed, during which indicators for measuring the performance of engineering facilities were identified from the open literature. Afterwards, some experienced facilities management (FM) practitioners were invited to attend a focus group meeting, which was intended to identify performance indicators which are applicable for evaluating engineering facilities in commercial buildings in Hong Kong; ascertain the feasibility of working out the applicable performance indicators; discuss any problems with recording or retrieval of the required data for deriving the applicable indicators. The processes and findings from such works are reported in the following.

2 Focus Group Discussion

At the initial phase of the study, altogether 71 indicators which are relevant to measuring the performance of engineering facilities were identified and they were grouped into five categories, namely physical (P), financial (F), task and equipment related (T), environmental (E), and health, safety and legal (H). A summary of the literature covering these indicators has been reported in a separate paper [8].

To determine the importance levels of the indicators and their feasibility in actual applications, a focus group meeting was convened in order to shortlist some key performance indicators (KPIs) from the 71 indicators. The meeting allowed certain flexibility in terms of its format and desired outcome. Fruitful data were

obtained from the direct interaction between the researchers and participants. In order to obtain as much useful findings as possible from a focus group discussion [9–11], the participants invited were highly experienced FM professionals.

A questionnaire, which was used to collect the required data during the focus group discussion, was designed with three sections. Section 1 contained questions on the personal particulars of the participants, including their genders, lengths of work experience, job titles, natures of their employers, and types of buildings/premises that they worked on. Section 2 consisted of two parts—Part 2A and Part 2B. Part 2A asked the participants to indicate on a five-point scale the importance levels of the listed indicators for evaluating the performance of engineering facilities in commercial buildings. Part 2B asked whether the performance indicators need to be included into the assessment scheme. The participants were also asked to provide the reasons for excluding the performance indicators. Section 3 sought for any other comments from the participants. Totally seven FM professionals participated in the meeting, which comprised the following sessions:

Session 1—Introduction

Each of the participants was given a set of handout for the introductory presentation, the definitions of the 71 performance inductors, and the questionnaire as mentioned above. Following a presentation on the background and purposes of the study and the focus group meeting, which was given by the meeting convenor, the participants were asked to complete Sect. 1 of the questionnaire.

Session 2—Brainstorming and discussion

The procedures below were taken for each of the listed performance indicators:

- The definition of each indicator was explained by the convenor.
- The importance level of each indicator was rated by the FM practitioners. Meanwhile, the FM practitioners were facilitated to discuss the feasibility and usefulness of the indicators in actual applications, and share their experiences in the use of those indicators, including any problems with recording or retrieval of the required data.
- The participants were further asked to vote whether the indicators need to be included into the assessment scheme. Only indicators supported by over half of the participants were shortlisted. For those indicators considered to be excluded from the scheme, the participants were asked to indicate the reasons behind.

Responses of the participants were recorded in Sect. 2 of the questionnaire. At the same time, the participants were free to brainstorm and suggest any other indicators which were not yet covered by those listed in the questionnaire.

Session 3—Review of the findings

After completing the whole discussion on all the indicators, the findings were reviewed collectively by the participants. Finally, they were allowed to fill in Sect. 3 of the questionnaire for any other comments they had.

3 Findings and Discussion

3.1 Background of Participants

All the focus group participants were male, with two being director of engineering, one chief engineer, two managers, and two assistant managers. Five of them were working for private companies and two of them were working for non-government public organizations. Having worked in the building industry for 14–31 years, they were highly experienced in managing the engineering facilities of various types of commercial premises such as office, retail, hotel, and restaurant.

3.2 Performance Indicators

Table 1 summarizes the descriptive statistics of the 71 indicators identified from literature and three additional indicators suggested by the focus group participants. The additional indicators were: ratio of total O&M cost to building income (F0), availability of fire services system (T28a), and availability of lift (T28b).

Ratio of total O&M cost to building income (F0) was suggested because, according to the discussion, it would be more meaningful to focus on the total cost in evaluating the performance of the facilities and all their O&M cost elements are covered by this indicator. Comparatively, it would be niggling to consider each individual cost element, such as percentage of personnel cost (F1), percentage of subcontractor cost (F2) and percentage of contractor cost (F3). In addition, it is not straightforward to figure out the costs of different kinds of manpower for determining F1 as some of the works may be done by contractors while some may be produced by in-house staff. F2 and F3, which are usually used for studying the proportion of works outsourced, would be useful for tracing the reasons for poor contractor performance.

Availability of fire services system (T28a) and availability of lift (T28b) were suggested because of the common safety concern of building end users. As far as availability of lift is concerned, the Electrical and Mechanical Services Department of the Hong Kong government has particularly specified that its monthly service availability should be not lower than 99 % [12]. This performance level is a reference for lift owners in setting specifications for the procurement of lift maintenance services. Instead of using the generic availability indicator (T28), adoption of T28a and T28b, which represent the performances of two relatively more critical facilities (i.e. fire services system and lift), is more specific and useful.

Table 1 Summary of the performance indicators and findings from the focus group discussion

Performance indicators		Mean importance level[a]	Rank	Important?[b]	To be included into the evaluation scheme?[c]
PI	Thermal comfort	4.29	7	Yes	Yes
P2	Visual comfort (e.g. illuminance and glare)	4.00	14	Yes	No
P3	Acoustic comfort (e.g. reverberation)	3.57	23	No	No
P4	Indoor air quality (e.g. total volatile organic compound, carbon dioxide (CO2) level, concentration of radon)	4.00	14	Yes	Yes
P5	Percentage users dissatisfied	4.14	11	Yes	Yes
P6	Number of users" complaints per year	3.14	26	No	No
FO	Ratio of total operation and maintenance (O&M) cost to building income	4.43	2	Yes	Yes
F1	Percentage of personnel cost	2.57	33	No	No
F2	Percentage of subcontractor cost	2.57	33	No	No
F3	Percentage of contractor cost	2.57	33	No	No
F4	Actual costs within budgeted costs [Excluding the extra expenditure for urgent or emergency works]	4.00	14	Yes	Yes
F5	Direct maintenance cost	2.14	52	No	No
F6	Breakdown severity [i.e. corrective maintenance cost /preventive maintenance cost]	3.14	26	No	No
F7	Equipment replacement value (ERV)	2.00	61	No	No
F8	Maintenance stock turnover	2.14	52	No	No
F9	Percentage of maintenance material cost	2.00	61	No	No
F10	Percentage of corrective maintenance cost	2.29	42	No	No
Fll	Percentage of preventive maintenance cost	2.29	42	No	No

(continued)

Table 1 (continued)

Performance indicators		Mean importance level[a]	Rank	Important?[b]	To be included into the evaluation scheme?[c]
F12	Percentage of condition based maintenance cost	2.71	30	No	No
F13	O&M cost per building area	4.43	2	Yes	Yes
F14	O&M cost per capacity of installation	2.29	42	No	No
F15	Cost of equipment added or replaced	1.86	68	No	No
F16	Energy expenditure per building area	2.14	52	No	No
F17	Energy expenditure per person	2.14	52	No	No
F18	Total safety and security expenditure	1.71	74	No	No
F19	Security expenditure per building area	1.86	68	No	No
F20	Security expenditure per person	1.86	68	No	No
F21	Building income per building area	1.86	68	No	No
F22	Total rentable value of the building	1.86	68	No	No
T1	Work request response rate	4.29	7	Yes	Yes
T2	Scheduling intensity	2.29	42	No	No
T3	Manpower utilization rate	2.29	42	No	No
T4	Manpower efficiency	2.29	42	No	No
T5	Manpower utilization index	2.29	42	No	No
T6	Preventive maintenance ratio (PMR)	2.71	30	No	No
17	Percentage of corrective (reactive) work	2.00	61	No	No
T8	Percentage of preventive (proactive) work	2.00	61	No	No
T9	Percentage of condition based maintenance work	2.00	61	No	No
T10	Percentage of improvement work	2.29	42	No	No
Til	Number of manhours per capacity of installation	2.29	42	No	No

(continued)

Table 1 (continued)

Performance indicators		Mean importance level[a]	Rank	Important?[b]	To be included into the evaluation scheme?[c]
T12	Number of completed work orders per staff	4.14	11	Yes	Yes
T13	Area maintained per maintenance staff	3.71	20	No	Yes
T14	Quality of scheduling	2.14	52	No	No
T15	Schedule realization rate	1.86	68	No	No
T16	Schedule compliance	2.71	30	No	No
T17	Work order turnover	2.14	52	No	No
T18	Backlog size	3.43	24	No	Yes
T19	Urgent repair request index (URI)	2.43	38	No	No
T20	Corrective maintenance time	2.14	52	No	No
T21	Preventive maintenance time	2.14	52	No	No
T22	Response time for maintenance	3.14	26	No	No
T23	Percentage compliance with required response time	2.14	52	No	No
T24	Number of maintenance induced interruptions	2.43	38	No	No
T25	Failure/breakdown frequency	4.29	7	Yes	Yes
T26	Meantime between failures (MTBF)	2.43	38	No	No
T27	Mean time to repair (MTTR)	2.57	33	No	No
T28	Availability	3.00	29	No	No
T28a	Availability of fire services system	4.43	2	Yes	Yes
T28b	Availability of lift	4.43	2	Yes	Yes
T29	Efficiency of facilities	3.71	20	No	No
T30	Gross floor area under safety and security patrol	2.00	61	No	No
El	Energy use index (EUI)	4.71	1	Yes	Yes
E2	Energy consumption per person	3.86	17	No	No
E3	Greenhouse gas emission per building area	4.43	2	Yes	Yes

(continued)

Table 1 (continued)

Performance indicators		Mean importance level[a]	Rank	Important?[b]	To be included into the evaluation scheme?[c]
E4	Conduction of energy audit	3.86	17	No	No
E5	Conduction of carbon audit	3.86	17	No	No
E6	Conduction of environmental assessment (e.g. LEED, BREEAM, BEAM Plus)	3.71	20	No	No
HI	Number of accidents per year	4.29	7	Yes	Yes
H2	Number of legal cases per year	2.57	33	No	No
H3	Number of compensation cases per year	2.29	42	No	No
H4	Amount of compensation paid per year	2.00	61	No	No
H5	Number of health and safety complaints per year	2.43	38	No	No
H6	Number of lost work days per year (i.e. sick leave day (s) given by doctor)	4.14	11	Yes	Yes
H7	Number of incidents of specific diseases in building per year (e.g. Legionnaire's disease)	3.29	25	No	No

[a] Scale for level of importance: 5 = very high; 4 = high; 3 = moderate; 2 = low; 1 = very low
[b] "Yes" for those with mean score not less than 4, else "No"
[c] Based on the participants' voting

3.3 Levels of Importance of the Indicators

The mean levels of importance of the indicators were calculated and ranked. Analysis of the descriptive statistics showed that the focus group participants considered 16 of the indicators as important - mean rated score being not less than 4 (out of a maximum score of 5). Such indicators include: five task and equipment related indicators (31.3 %), four physical indicators (25.0 %), three financial indicators (18.8 %), two environmental indicators (12.5 %), and two health, safety and legal indicators (12.5 %). Among these 16 indicators, the top-ranked indicator, with the highest mean score, was energy use index (EUI), i.e. E1. This is probably

due to the fact that energy use, especially electricity consumption, is a substantial cost burden in operating engineering facilities of commercial buildings. In fact, about 60 % of the total electricity in Hong Kong was used in the commercial sector and the amount of such consumption has been increasing continuously [13]. Therefore, using indicator E1 can help identify measures for curbing the use of electricity in commercial buildings.

Several indicators recorded the same mean score and were ranked second in the list. They were: O&M cost per building area (F13), availability of fire services system (T28a), availability of lift (T28b) and greenhouse gas emission per building area (E3).

For indicator F13, it is important to know the amount of money spent on operating and maintaining the facilities as the engineering department has to prepare O&M budgets and monitor their actual expenditures. If the O&M cost can be controlled and minimized, that will be a reflection of the effort made by the FM team in managing the performance of the facilities. Logging O&M cost and normalizing it by building area (F13), moreover, can facilitate comparisons and performance benchmarking among peer buildings.

As mentioned earlier, availability of fire services system (T28a) and availability of lift (T28b) were important performance indicators of safety in buildings. As for environmental sustainability, greenhouse gas emission per building area (E3) is important and carbon audit has in recent years become a hot topic in the building industry. In 2012, the Hong Kong government has enacted the Building Energy Efficiency Ordinance (Cap. 610) to stipulate the mandatory conduction of energy audits for commercial buildings [14], which is a significant step after the launch of the Buildings Energy Efficiency Funding Schemes in 2009 to subsidize building owners to conduct energy-cum-carbon audits and energy improvement projects [15].

3.4 Shortlisted Performance Indicators

Of the 16 important indicators, 15 were shortlisted for inclusion into the scheme. They were: thermal comfort (P1), indoor air quality (P4), percentage users dissatisfied (P5), ratio of total O&M cost to building income (F0), actual costs within budgeted costs (F4), O&M cost per building area (F13), work request response rate (T1), number of completed work orders per staff (T12), failure/breakdown frequency (T25), availability of fire services system (T28a), availability of lift (T28b), energy use index (EUI) (E1), greenhouse gas emission per building area (E3), number of accidents per year (H1), and number of lost work days per year (H6). These indicators were classified as “Cat I - Important and need to be included” (see Table 2).

Visual comfort (P2) was also an important performance indicator but, after deliberations at the focus group meeting, it was excluded from the scheme. In other words, it belonged to “Cat II - Important but need to be excluded”. The reasons for

Table 2 Categories of performance indicators

	Important		Less important	
Need to be included	Cat I		Cat III	
	P1	Thermal comfort	T13	Area maintained per maintenance staff
	P4	Indoor air quality	T18	Backlog size
	P5	Percentage users dissatisfied		
	F0	Ratio of total O&M cost to building income		
	F4	Actual costs within budgeted costs		
	F13	O&M cost per building area		
	T1	Work request response rate		
	T12	Number of completed work orders per staff		
	T25	Failure/breakdown frequency		
	T28a	Availability of fire services system		
	T28b	Availability of lift		
	E1	Energy use index (EUI)		
	E3	Greenhouse gas emission per building area		
	H1	Number of accidents per year		
	H6	Number of lost work days per year		
Need to be excluded	Cat II		Cat IV	
	P2	Visual comfort	The remaining 56 performance indicators.	

this decision included the difficulties in conducting the necessary measurements for working out this indicator, e.g. measurement of glare at different times in various building areas. The resources needed for undertaking such measurement, according to the experience of the participants, would be far more than the benefit that may be obtained from including such an indicator in the assessment scheme.

Two other indicators, namely area maintained per maintenance staff (T13) and backlog size (T18), were considered as comparatively less important but were recognized as useful indicators under the performance assessment scheme. Indicator T13 would enable the FM team to check if the manpower (number of staff, e.g. technicians) is sufficient to carry out daily O&M works. This indicator would be a useful reference for the team to request for more resources in case the manpower is found insufficient. Backlog size (T18), which shows the percentage of overdue work orders, is useful for indicating the work efficiency of the FM team. Both T13 and T18 were classified as "Cat III - Less important but need to be included".

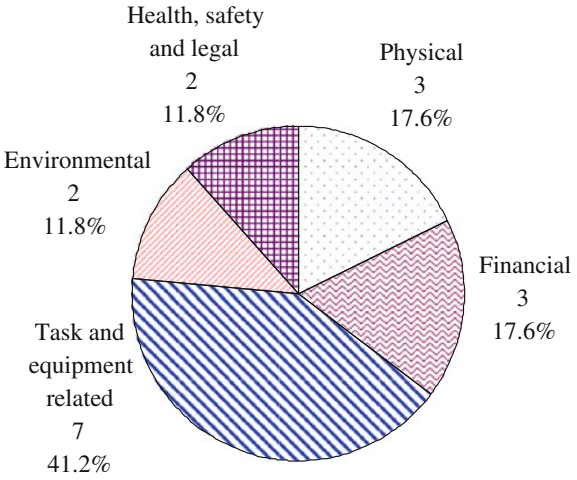


Fig. 1 Distribution of the various categories of shortlisted indicators

The remaining indicators, which are less important, were dropped out from the scheme, i.e. they were grouped under "Cat IV—Less important but need to be excluded". The predominant, common reason for their exclusion from the scheme was that it would be too time consuming to collect the necessary data for finding out such less-important indicators.

Eventually, 17 key performance indicators (KPIs), belonging to Cat I and Cat III, were shortlisted for inclusion in the assessment scheme. Among them, seven (41.2 %; see Fig. 1) were task and equipment related indicators. There were only two to three KPIs coming from each of the remaining four groups – physical; financial; environmental; and health, safety and legal. This shows that, in terms of quantity, task and equipment related indicators would be dominant in the assessment scheme.

4 Conclusion and Future Works

A focus group meeting, involving in-depth discussions and exchange of expert opinions among seven FM professionals in Hong Kong, was conducted to determine the applicability and importance of 71 performance indicators which were identified from relevant literature. Three additional performance indicators were suggested by the focus group participants. After deliberation, 16 out of all the 74 indicators were considered to be important, among them 15 were considered to be essential for inclusion into the assessment scheme. The common reason for those excluded from the scheme was also identified. In addition to the 15 indicators, two other indicators (area maintained per maintenance staff, and backlog size), in view of their usefulness in performance evaluation, were added into the scheme. The

majority of the 17 KPIs, in terms of quantity, were dominated by task and equipment related indicators.

Despite the large volume of preparation and organization works required for convening the focus group meeting, it proved to be an effective means for identifying useful KPIs among the long list of usable performance indicators. To further single out the most useful KPIs, a large-scale questionnaire survey has been designed to collect the opinions of more FM professionals. When the survey findings are available, the most useful KPIs will be analysed further, for example, by an analytic hierarchy process [16], to find out their importance weights. With the latter determined, an assessment scheme will be established for use in evaluating the performance of engineering facilities in commercial buildings in Hong Kong.

Acknowledgement The authors are grateful to the support given by the Building Services Operation and Maintenance Executives Society (www.bsomes.org.hk) to the focus group meeting.

References

1. Census and Statistics Department (2013) Population. Census and statistics department. http://www.censtatd.gov.hk/hkstat/sub/so20.jsp. Accessed 29 March 2013
2. The Government of the Hong Kong Special Administrative Region (2013) Hong Kong—the Facts, Geography. The Government of the Hong Kong Special Administrative Region. http://www.gov.hk/en/about/abouthk/facts.htm. Accessed 29 March 2013
3. Hong Kong Tourism Board (2013) Overnight visitors to Hong Kong up 6.5 % in 2012. Hong Kong Tourism Board. http://partnernet.hktb.com/filemanager/pressrelease/Tourism%20Stat%2012%202012.pdf. Accessed 29 March 2013
4. Lai JHK, Yik FWH, Jones P (2008) Expenditure on operation and maintenance service and rental income of commercial buildings. Facilities 26(5/6):242–265
5. Lai JHK (2010) Operation and maintenance budgeting for commercial buildings in Hong Kong. Constr Manag Econ 28(April):415–427
6. Thor CG (1991) Performance measurement in a research organization. Natl Prod Rev 10 (4):499–507
7. Webster C, Hung L (1994) Measuring service quality and promoting decentring. TQM Mag 6 (5):50–55
8. Man CS, Lai JHK, Yik FWH (2013) Developing a research framework for studying performance evaluation of engineering facilities in commercial buildings in Hong Kong. In: Proceedings of the 19th International CIB World Building Congress. Brisbane, 5–9 May 2013
9. Berg BL (2009) Qualitative research methods for the social sciences, 7th edn. Allyn & Bacon, Boston
10. Fern EF (2001) Advanced focus group research. Sage Publications, Thousand Oaks, California
11. Hesse-Biber SN, Leavy P (2004) Approaches to qualitative research – a reader on theory and practice. Oxford University Press, New York
12. Electrical and Mechanical Services Department (2009) Particular specification for lift and escalator installations. Electrical and Mechanical Services Department, Hong Kong
13. Electrical and Mechanical Services Department (2012) Hong Kong energy end-use data 2012. Electrical and Mechanical Services Department, Hong Kong
14. Department of Justice (2012) Buildings Energy Efficiency Ordinance—Energy audit requirement. Department of Justice. http://www.legislation.gov.hk/blis_ind.nsf/CurAllEngDoc/2E376A4736C46EB548257814001089AD?OpenDocument. Accessed 29 March 2013

15. Environment and Conservation Fund (2012) Buildings energy efficiency funding schemes. Environment and Conservation Fund. http://www.building-energy-funds.gov.hk/en/news/index.html. Accessed 30 March 2013
16. Saaty TL (1980) The Analytic hierarchy process. McGraw-Hill, New York

Asset Management Decisions—Based on System Thinking and Data Analysis

Helena Kortelainen, Susanna Kunttu, Pasi Valkokari and Toni Ahonen

Abstract Asset related data is collected in several information systems (e.g. enterprise resource management (ERP) and computerized maintenance management systems (CMMS) systems) at industrial plants. Information systems including asset related data are typically used for operational level decisions (e.g. creating maintenance work orders) but maintenance history data is also valuable when making asset management level decisions (e.g. investment decisions). Even though there is a huge amount of stored data, tacit knowledge is needed for risk conscious asset decisions both for supplementing the data contained in IT-systems and for creating the understanding of the production system itself and its interrelationships. The paper describes how data collected from ERP and CMMS system can be utilized when improving operational efficiency and researching investment opportunities and evaluating investment options.

1 Introduction

Asset related data is collected in several information systems (e.g. enterprise resource management (ERP) and computerized maintenance management systems (CMMS) systems) at industrial plants. In addition, asset data is collected, among other repositories, in control room diaries and condition monitoring systems. The recent survey of EFNMS [1] reveals that even though in some companies 70–95 % of asset events are recorded a major part of companies has significantly lower

H. Kortelainen (✉) · S. Kunttu · P. Valkokari · T. Ahonen
Technical Research Centre of Finland, P.B. Box 1000, 02044 VTT, Finland
e-mail: helena.kortelainen@vtt.fi

S. Kunttu
e-mail: susanna.kunttu@vtt.fi

P. Valkokari
e-mail: pasi.valkokari@vtt.fi

T. Ahonen
e-mail: toni.ahonen@vtt.fi

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_92

portion of registered events. The quality of the data stored in the IT systems has been recognized as a major challenge in several industries (e.g. [2]). Often the collected data is limited to inventory data only [3]. However, the main problems seem to be organisational issues like inefficient organisational roles, missing placement of responsibilities and lack of training rather than issues dealing with IT systems themselves [4]. As data are created and used daily in all industrial operations poor quality of data and incomplete information obviously can have negative effects on company performance.

Even though there is a huge amount of stored data, tacit knowledge is needed for risk conscious asset decisions both for supplementing the data contained in IT-systems and for creating the understanding of the production system itself and its interrelationships. In the capital-intensive industries production assets typically have long life cycles and major changes may occur in the external business environment. This may lead to a challenge of ever shortening market life of products and relatively long life cycle of production systems [5, 6]. Thus, understanding the constraints given by the production technology and uncertainty associated to the market life of products and other external factors is crucial to any effective strategy [7]. The history data alone is not sufficient basis for future forecasts but tacit knowledge, expert judgment and other data sources have to be utilized. There seems to be lack of adequate systematic approaches to combine different data sources when supporting asset decisions of different time horizons.

Information management can be placed into the list of the issues having room for improvement and there is a need to improve both the data quality and handling. However, the data analysis and utilisation aspect is equally important. There is a need for business processes that enable the analysis to help define asset strategies based on corporate and/or department goals [8] or business value creation [9]. A crucial issue for business success is also the alignment of asset strategy with the company´s strategic objectives [10, 11] and performance indicators are a part of this process. Information systems including asset related data are typically used for operational level decisions (e.g. creating maintenance work orders). However, the history data containing the information on the physical asset events, their upkeep and modifications is crucially important for systematic maintenance planning and valuable contribution when making tactical or strategic level decisions. This paper draws on literature, surveys, case studies and consulting projects to describe how data collected in ERP and CMMS system can support asset related decision ranging over a wide time horizon.

2 Asset Data Utilisation for Decision Making of Different Time Horizons

Wide and versatile utilisation of the collected data motivate for developing data collection both in terms of quality and amount. The asset related data stored in information systems are typically used for operational level decisions in creating maintenance work orders and recording the work done but the data has much wider

untapped exploitation potential. The following sections collect examples from our case studies that illustrate the data analysis and utilisation in maintenance planning, criticality assessment and investment portfolio planning.

2.1 System Approach for Maintenance Planning and Performance Improvement

Many ERP systems can only look for an asset as a unit and in many industrial companies the data is recorded in the CMMS on a plant level or to a production line —even if the exact allocation would be possible. One of the key challenges in evaluating the asset condition is the ability to drill down to the necessary level of detail in the asset hierarchy [12]. Hierarchical analysis requires that event data is recorded to the item level when different system levels can be analysed by combining data and breaking down. The top-down approach offers clear benefit because it helps to identify those issues that are the most significant to the overall operational efficiency of the plant [13, 14]. Figure 3 represents an example of a hierarchical analysis from the food industry. The break-down of the maintenance costs that are first allocated to the production lines and further to subsystems and items. In this case example [15] production line 1 and especially the subsystem 1 caused a major part of the maintenance costs. In this case the driver for analysing the system was high maintenance costs and it might have been possible to identify cost drivers also from the item level data. Top-down approach helps also to understand the causal relationships and to identify wider problem areas which enable more holistic solutions from system point of view. Such wider problems may remain unidentified if the analysis is carried out on the equipment level only (Fig. 1).

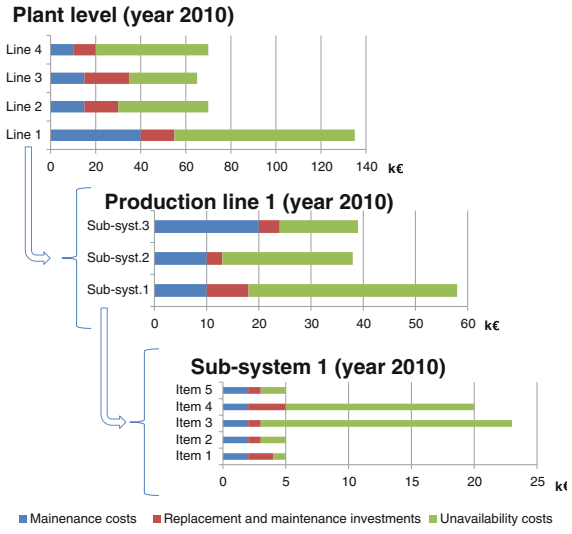


Fig. 1 Analysing the availability performance data (modified from [15])

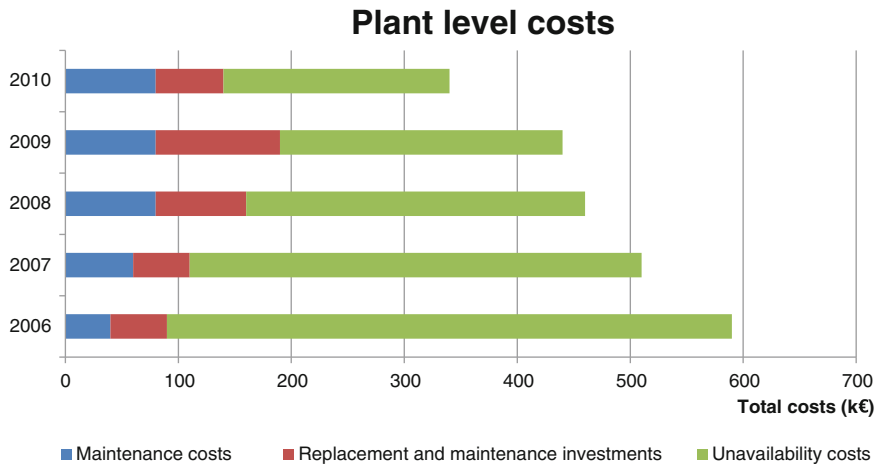


Fig. 2 Impact investment on maintenance on plant level costs (modified from [15])

Unavailability costs are bound with uncertainty due to the changes in product demand and price. For this reason unavailability costs are often omitted and the analysis is based on direct maintenance costs. In our studies, the average price of lost production per hour describes the average situation. Even though this is not exactly correct the order of magnitude is right and this information is adequate for identifying maintenance development needs and for planning purposes.

Yearly maintenance costs are often used as a measure of successful maintenance function. This measure is unbalanced as the short-term reduction of maintenance leads to declining condition of production assets and causes production losses in the long term. Taking the unavailability costs into consideration and utilising the recorded history data the development of maintenance function may reveal to be a profitable investment. Figure 2 shows an example how increase in maintenance costs can decrease total costs. In this example from the food industry, the replacement and modernisation investments together with an increase in maintenance efforts have reduced the total costs by 40 % in five years.

The overall objective of any production plant is to maximise economic outcome which usually means that also the operational costs have to be minimised. In this context it is important that the economic consequences of production losses are taken into consideration in the decision making process. The minimisation of the overall costs is not successful only by reducing individual cost items but by choosing the most profitable investments. This approach contributes also to business continuity and sustainable operation.

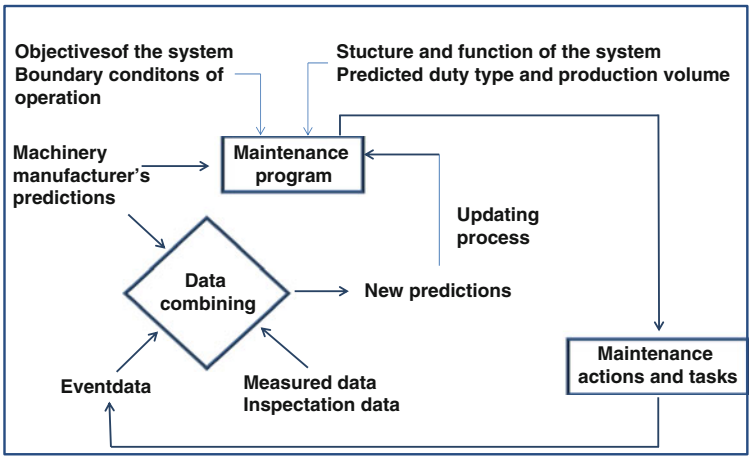


Fig. 3 Schematic presentation of a continuously improved maintenance program

2.2 Updating the Maintenance Programme

According to the IEC-60300-3-11 standard [16] the maintenance programmes are composed of the initial programme and an on-going, dynamic programme. The initial programme of maintenance recommendations are often delivered by the manufacturer. The standard does not present detailed guidelines how to develop and update the programme. If event data is collected in CMMS or a corresponding information system, the history data could be used as sketched in the Fig. 3.

Event data seldom alone contains enough information for necessary updates and adjustments as system structure, function and/or boundary conditions may have changed. In addition maintenance programme should respond to the predicted future duty type and conditions of the system. RCM (Reliability Centered Maintenance) (see e.g. [16]) in its different forms, like Value-driven maintenance planning [9] offers a systematic method that takes the objectives of the plant as the reference point for specifying functional requirements for the equipment locations and equipment. An alternative process is based on carrying out FMECA (Failure Modes, Effects and Criticality Analysis, see e.g. [17]) to produce information on failures that have not been covered by the initial or current programme. The FMECA analysis can be supplemented by RCM decision logic tree to develop corresponding preventive maintenance tasks [14]. All these analysis methods utilise available recorded maintenance data but draw also from systematic expert elicitation process.

One example of the criticality assessment where the number of failures is based on event data is shown in Fig. 4. In this case, expert judgement is utilised to evaluate the cost of different failure types. With scarce history data, and especially with rare failure events, a criticality analysis based on event data may be incomplete and more representative assessment suggests the incorporation of FMECA analysis [14, 18].

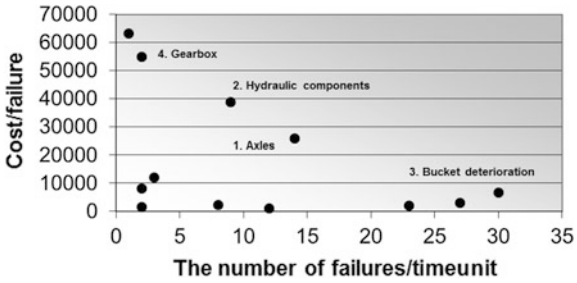


Fig. 4 Criticality assessment of components and subsystems based on event data in electricity transmission system of a mobile working machine [14]

The criticality may arise from a high cost of an individual failure (e.g. gearbox) but also from very frequent failures (e.g. bucket deterioration). In the latter case options may include improvements in the maintenance program but also re-design, replacement or modernisation.

2.3 System Approach for Evaluation of Investment Portfolio

There is a general agreement that investment decisions are the most important decisions made by corporations. Especially replacement, modernisation and maintenance investments benefit from history data contained in ERP and/or CMMS. As the investment is aimed to create future competitive advantage the evaluation of options cannot be based on mere continuum of the historical development. A well-structured approach for collecting expert views on market position, technological constraints and company-specific factors is required [7, 11, 19].

Typical investment decision problems require decision makers to select a subset of available decision options that return a maximum profit or fulfil some other performance objectives. At the same time, the overall risk of a portfolio needs to be controlled [20]. System approach is required in order to avoid sub-optimisation.

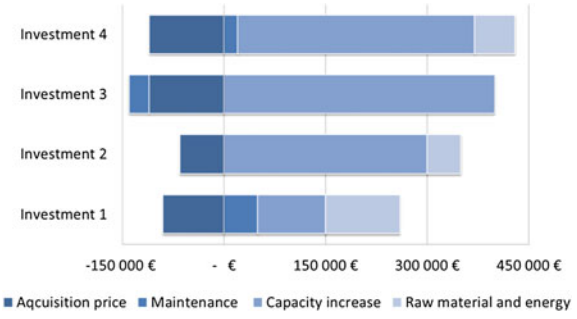


Fig. 5 The graph presenting the structure of the costs and profits of various investment options

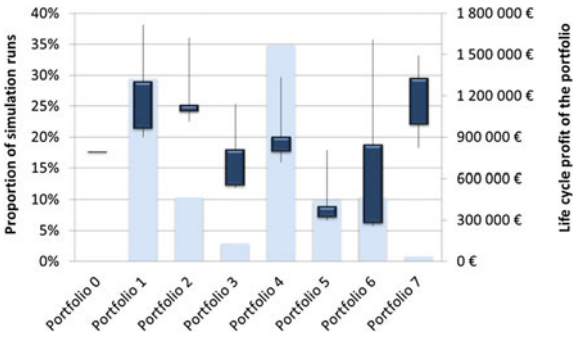


Fig. 6 Results of the MonteCarlo simulation for investment portfolios indicate the profitability and expected variation of each investment portfolio

Our approach for investment portfolio evaluation [7, 19] offers evaluation methods for optimizing and supporting business-driven asset strategy decisions while taking into account economic constraints. The approach utilises systematic methodologies like investment appraisal, AHP (Analytic Hierarchy Process), expert judgement, risk analysis and MonteCarlo simulation. Figures 5 and 6 illustrate results of a study carried out in process industry. In our approach, the evaluation of investment options proceeds on two levels: individual investment option and portfolio level. Each portfolio consists of a set of individual investment options.

Different evaluation techniques incorporate uncertainty in different ways. In our approach sensitivity analysis was conducted by Monte Carlo simulation in order to find out which factors are the most relevant in terms of risk regarding the investment profitability. Expert judgement is used to define some technical and economic values if historic data from different information systems (ERP, CMMS etc.) is not available. Data resulting from statistical data analysis, trend analysis and expert judgement is used in the evaluation model as input values when evaluating investment and investment portfolios.

The structured evaluation process and visual reports help the decision maker to select the most profitable investment portfolio that offers the best response to the likely changes in the market and risk attitude.

3 Discussion

In the previous sections examples on using the history data drawn from the IT systems like CMMS and ERP were presented. For industrial systems, the component or item level failure rates are typically low. Thus mere history data does not contain all the relevant information about factors influencing the failure tendency. In most examples event data was supplemented by expert judgement for two reasons: firstly, the data contained in IT systems is often incomplete or of poor quality

Table 1 Asset management related tasks and data sources in the context of planning time horizon

Time horizon	Day-to-day operation		Planning for mid and long term		Future scenarios
Organisational level	Operational level (on demand)	Operational level (planned)	Tactical/managerial level	Strategic level	Visions
Typical data	Online sensors and measurements, inspection, Production data	Sensor data, event data, inspections	Event data, expert data	Forecasts and expert data, asset history data	Forecasts, scenarios,
Typical tasks	Carrying out break down maintenance and opportunity based maintenance	Carrying out planned and scheduled maintenance	Develop producion and maintenance strategies, Minor modernisation and replacement investments	Strategic decisions e.g. outsourcing, investments in production assets, and IT, training & education,	Market strategies, capacity Strategies, M&A actions, Greenfield investments
	Carrying out condition based maintenance				

and secondly, the analysis requires insights of future development. The time horizon of asset decisions covers day-to-day decisions on the shop floor to the strategic decisions dealing with development of the production capacity with Greenfield investments or merger or acquisition. The wide time range is also reflected to the data requirements.

As illustrated in the Table 1, asset related data contains a wide range of knowledge from explicit data that is captured by manual or computerised systems to tacit knowledge and insights that are delivered by individual workers, experts and executives. The use of expert judgement also includes challenges such as subjectivity and overconfidence of the experts. There is also a danger that someone dominates the expert session and the outcome is deficient or even biased. Thus, the coverage of the expert group and competence of the experts are of prime importance. The task is then to choose suitable elicitation methods and approaches to grasp this knowledge, means to make this knowledge visible and to connect the expert opinions to the explicit and history data.

An important issue that was not dealt with in this paper is the use of sensor data provided by automation and condition monitoring systems, or the data collected from the inspections. There is a wide body of literature dealing with condition based monitoring and risk-based inspections [21]. Measurement and inspection data could be useful when updating the maintenance programme as sketched in the Fig. 3. From the production availability performance point of view the unavailability costs due to the planned and unplanned maintenance actions are the same (even though labour and other costs are different). Opportunity-based maintenance aims at using the stoppages due to the production reasons for carrying out maintenance tasks. Sensor and measurement data together with data analysis and visually presented interpretation of the data [22] could help to adjust the maintenance tasks to the schedule of production. Also, the emerging trend to allocate part of the maintenance tasks to operators instead of maintenance technicians contribute to this.

Asset management is often used as a synonym for maintenance management. Physical asset management deals with costs, benefits and risks. European Federation of National Maintenance Societies [23] defined asset management as “the optimal life cycle management of physical assets to sustainably achieve the stated business objectives”. This definition implies a system approach and presupposes system thinking capabilities when developing tools for asset related decision making.

4 Conclusions

EFMNS survey [23] poses data as one of the key areas of development in spite of the fact that there is a huge amount of stored data in company IT systems. The data is not fully exploited in the asset related decision making – partly because the data is of poor quality and partly because there is a lack of practical systematic approaches to support asset decisions of different time horizons. As several decisions deal with future development tacit knowledge and forecasts are needed for

supplementing the data contained in IT-systems and for creating the understanding of the production system itself and its interrelationships.

This paper presented some examples of analysis that have system thinking as a starting point and describe how data collected in ERP and CMMS system can support asset related decision ranging over a wide time horizon. The presented examples deal with maintenance planning, maintenance cost breakdown, use of FMECA to gather expert knowledge, criticality assessment and investment portfolio planning. These examples cover the mid and long term planning and the presented approaches deal with managerial and strategic decision making.

The examples highlighted that the event data has to be recorded on an item level. Thus those items causing high maintenance costs are identified and the data can also be used when evaluating the improvement options. The applied IT system or application software should allow to "drill down" from the system to item level to first find wider problem areas and then define specific targets for concrete development tasks. The top-down view is important as it helps to analyse an item in a system context.

Often the history data has to be supplemented with expert judgement and this is done case by case. It is necessary to create capabilities to combine data and knowledge from different sources like expert knowledge with event data and measurement and inspection data with event data in order to overcome the limitations of data quality and coverage. Recording the expert data into the IT-systems for future use is not a standard property of a CMMS or ERP-system, neither are the tools to combine data from different sources.

It is also of prime importance to collect data on unavailability time due to the planned or unplanned shutdowns. This underlines the impact of maintenance and shows the value-adding capability of maintenance tasks. If the focus is in the direct maintenance cost only the understandable aim to reduce cost may turn detrimental to the production availability performance in the long run. Without properly collected data the cost – benefit relationship and the value of maintenance is not possible to indicate.

Finally, many asset related decisions require insights of future development. Typical investment decision problems require decision makers to select a subset of available decision options that return a maximum benefit over a time horizon that may be decades. This is not possible with the data contained in the IT systems alone even though the data may offer a lot of basic knowledge of the existing systems. There is still a lot of space for the development of forecasting methods, uncertainty management and systematic approaches for optimizing and supporting business-driven asset strategy decisions.

References

1. EFNMS (2012) How organisations manage their physical assets in practice. EFNMS asset management survey 2011. European Asset Management Committee within EFNMS. European Federation of National Maintenance Societies vzw

2. Silvola R, Jaaskelainen O, Kropsu-Vehkapera H, Haapasalo H (2011) Managing one master data—challenges and preconditions. Comm Assoc Inf Syst 4(23):63–72
3. Kunttu S, Kiiveri J (2012) Take the Advantage of dependability data. Maintworld 4(3):24–26
4. Haug A, Arlbjørn J, Zachriassen F, Schlichter J (2013) Master data quality barriers: an empirical investigation. Ind Manag Data Syst 113(2):234–248
5. Tranfield D, Denyer D, Burr M (2004) A framework for the strategic management of long term assets (SMoLTA). Manag Decis 42(2):277–291
6. Teece DJ, Pisano G, Shuen A (1997) Dynamic capabilities and strategic management. Strateg Manag J 18(7):509–533
7. Kortelainen H, Räikkönen M, Komonen K (2013) Corporate asset management—a semi-quantitative business—driven approach for evaluating improvement options. MPMM 2013 maintenance performance measurement and management conference. Sept. 12–13 2013. Lappeenranta, Finland
8. McNeeney A (2005) Improve asset performance management. Hydrocarbon processing 84 (12):61–67
9. Rosqvist T, Laakso K, Reunanen M (2009) Value-driven maintenance planning for a production plant. Reliab Eng Syst Safety 94(1):97–110
10. Parida A (2012) Asset performance assessment. In: Van der Lei T, Herder P, Wijnia Y (eds) Asset management, 101–113. Springer Science, Business Media B.V
11. Komonen K, Kortelainen H, Räikkönen M (2012) Corporate asset management for industrial companies: an integrated business-driven approach. In: Van der Lei T, Herder P, Wijnia Y (eds) Asset management, 67–86. Springer Science, Business Media B.V
12. Melvin G (2012) Top 5 questions you should ask about your assets and how pas55 can help you answer them. MaintWord 4:24–27
13. Kortelainen H, Pursio S (2001) Availablity performance stands for plant efficiency. Paperi ja Puu Paper and Timber 84(2):292–296
14. Ahonen T, Reunanen M, Heikkilä J (2006) Updating a maintenance programme based on various information sources. In: Konbin 4th International conference on safety and reliability. Cracow, Poland. May 30 – Jun 2. Air Force Institute of Technology. Poland, Warsaw. 6 p
15. Valkokari P, Kunttu S, Ahonen T (2011) Maintenance data in productive maintenance. Promaint 25(2):24–27 In Finnish
16. IEC 60300-3-11. Dependability management—Part 3–11: Application guide. Reliability centered maintenance. In: International electrochemical commission IEC. 90 p
17. IEC 60300-3-9. Dependability management—Part 3: Application guide. Section 9: Risk analysis of technological systems. In: International electrochemical commission IEC. 47 p
18. Kunttu S, Kortelainen H (2004) Supporting maintenance decisions with event and expert data. In: Proceedings of the annual reliability and maintainability symposium 2004. Los Angeles, CA, 26–29 Jan. 2004. IEEE, pp 593–599
19. Heikkilä A, Komonen K, Räikkönen M, Kunttu S (2012) Empirical experiences of investment portfolio management in a capital-intensive business environment. Int J Strat Eng Asset Manag 1(2):117–134
20. Better M, Glover F (2006) Selecting project portfolios by optimizing simulations. Eng Econ 51(2):81–97
21. Holmberg K, Adgar A, Aitor A, Jantunen E, Mascolo J, Mekid S (eds) (2010) E-Maintenance. Springer, London
22. Mikkonen H, Markkanen J (2013) Improved production and maintenance efficiency by operator driven reliability (ODR). COMADEM 2013. In: International Conference on Condition Monitoring and Maintenance. 11–13 June. Helsinki, Finland
23. EFNMS (2009) A definition of Asset Management. Minutes of the meeting. European federation of national maintenance societies, Trondheim, Norway

Executing Sustainable Business in Practice—A Case Study on How to Support Sustainable Investment Decisions

Susanna Kunttu, Markku Reunanen, Juha Raukola, Kari Frankenhaeuser and Jaana Frankenhaeuser

Abstract Too often, B2B negotiations only look at the purchasing cost and do not take the costs and effects on the environment of the whole life cycle of the product into account. Products that have less impact on the environment often have a higher purchasing price. It is generally believed that when a customer can see the estimates of the use period costs, he can accept the higher purchasing price more easily. This paper describes a practical LCC tool developed for life cycle cost calculations and how this kind of tool can be used by manufacturers in developing their own product portfolios and selling their products. For manufacturers, LCC calculations can reveal weak points of their solutions. In negotiations with customers, the LCC calculation indicates the kind of value they will obtain in the long run and guides them to choose products that cause less harm to the environment and have financial benefits.

1 Introduction

Industrial solution providers have seen that even though it has been argued in public discussions that sustainability and life cycle management are important issues, the practical actions are often still quite marginal. In investment decisions, the overriding criterion tends to be the investment cost. This is understandable as practical tools to analyse and systematically compare life cycle costs of different solutions have not been widely available and, thus, the acquisition price is the cost indicator whose value is most easily available. The challenge of practical tools is that the cost calculation is highly dependent on the case, as the cost structures are different.

S. Kunttu (✉) · M. Reunanen
Risk and Reliability Management, Technical Research Centre of Finland, Tampere, Finland
e-mail: Susanna.Kunttu@vtt.fi

J. Raukola · K. Frankenhaeuser · J. Frankenhaeuser
Elcon Solutions Oy, Paimio, Finland

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_93

Hardin [1] has argued that people are tempted to make decisions that serve their own interests, even if they are known to have a negative effect on the common interest. Hardin has presented the concept 'the tragedy of the commons' describing the dilemma of own interest versus common interest. This concept explains why one of the three pillars of sustainability is economic. It is hard to see a solution that does not provide economic benefits for its producers and users when it is on the market. The other two pillars of sustainability are environmental and social.

Gluch and Baumann [2] have criticized the use of the LCC calculation in environmental decision-making because its focus is not on the environment but the economy. They are of course right in this. When considering sustainability and its three pillars, sustainable decision-making is seen as multi-criteria decision-making, which justifies the use of the LCC calculation as a tool. In this case, the environmental and economic goals are close together, e.g. energy savings or number of replacements. As no data were available to conduct a life cycle assessment that considers the environmental impact in this case, it was decided to focus on the LCC approach.

To bridge the gap between practical decision-making and visions about sustainable decisions, a practical tool was developed to calculate the life cycle cost. The aim of this paper is to present how the tool was created and how a solution provider can use the life cycle cost calculation to support its customer's investment decision-making and own product portfolio management. In general, investment decisions are supported by profit comparison methods that consider the costs and revenues of the investment [3]. In this case, the product considered did not provide direct profits, but it could reduce life cycle costs. Thus, a cost comparison method was selected for use in this case.

The remainder of this paper has five sections. The paper first presents the research hypothesis and methodology used. The third section describes the case company and the product considered in this research. The fourth section describes how, in practice, the LCC calculations were conducted in this case. The fifth section presents the use of the LCC calculation. The last section discusses the results of this study.

2 Research Hypothesis and Methodology

The main research hypothesis in this study is that 'A solution provider can support its customer's sustainable investment decisions by providing the customer with detailed life cycle cost information.' The questions related to the research hypothesis concern what kind of a tool is needed to conduct the LCC calculations and whether customers should perform the calculations themselves or be supported by the solution provider.

The research method applied was a qualitative case study. The currently available research data were collected in workshops with the case company and through structured interviews with the customers. In order to test the hypothesis, the case-

specific LCC calculation details and related prototype tool to perform the calculations were first developed in close co-operation with the case company.

The case-specific calculation method and related tool that were developed are being tested in real customer cases. The results of the tests are collected from customers through structured interviews conducted straight after the test situation. The tests with customers are still ongoing as it has been difficult to schedule interviews on busy calendars, even though the customers are very interested in testing the calculation tool.

3 Case Description

The case company is a small company providing power supply systems to the energy, ICT, transport and process industries. The battery back-up systems are necessary to guarantee 24/7 operation of critical devices also in any failure situations of the electrical mains network. Battery back-up DC power supply system solutions are being used in many power plants and stations, substations and other locations, including e.g. uninterrupted power supply of process automation. The case company's products are typically customized solutions for its B-to-B customers who are project suppliers of larger systems and integrate the solutions delivered by the case company into their own offerings to end-users.

With regard to the case company's network position in the upstream direction there are large component and equipment suppliers, as well as network partners participating in assembly, manufacturing or R&D. Downstream, there are B-to-B customers and end-users from several sectors [4].

The product considered in this case is an uninterrupted DC power solution (see Fig. 1). DC power solutions are advanced battery back-up power supply systems. They guarantee uninterrupted operation of customer's large utility and industrial equipment and meet the power requirements of various system applications efficiently and reliably. Due to the modular structure of the systems, they can be customized according to the use requirements. The DC power systems are designed using advanced switched mode rectifier modules and can be completely setup, controlled, and monitored locally or remotely through network. Clean/renewable energy sources can be utilised in addition to conventional ones. The lifetime of DC power systems is typically about 20 years and their acquisition price is rather low in proportion to the lifetime costs, which makes the life cycle cost a better indicator for decision-making.

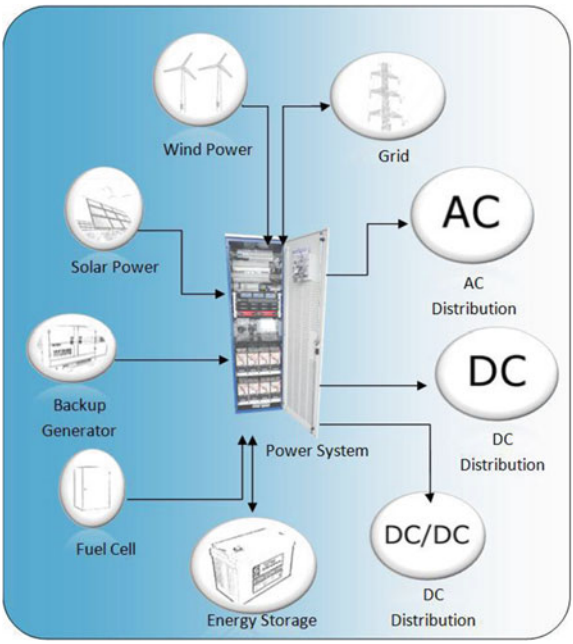


Fig. 1 Uninterrupted DC power solution

4 LCC Calculation

The first task of this research was to establish the life cycle cost calculation for the current case. This was done according to the guidelines given in the dependability management standard [5]. The tasks to conduct the LCC calculation listed in the standard are:

- cost breakdown structure
- product/work breakdown structure
- selection of cost categories
- selection of cost elements
- estimation of costs
- presentation of results

When applicable, they may also include:

- environmental and safety aspects
- uncertainty and risks
- sensitivity analysis to identify cost drivers

By applying the standard method, a systematic approach and comprehensive LCC analysis were assured.

Standards can generally only provide guidelines for actions to achieve strategic goals and cannot include elaborated instructions, as, in practice, actions are heavily

dependent on the company and its operational environment. The next chapters describe how LCC analysis has been performed in practice for a power supply system following the guidelines given by the standard.

4.1 Cost Breakdown Structure

The life cycle cost calculation of a system is based on a cost breakdown structure that divides the total life cycle cost into relevant cost categories and further into concrete cost parameters that are much easier to estimate than the total cost. The dependability management standard [5] mentions four tasks concerning identification of costs. In this case, those tasks were combined into one task: the creation of the cost breakdown structure. In this case, a hierarchical top-down approach was applied, providing a systematic method to create a comprehensive cost breakdown structure. The highest level of the hierarchy was a life cycle phase and the lowest level was formed by the cost parameters for which numerical values could be given. The depth of the hierarchy can vary in different branches depending on the data availability. For example, the acquisition price can be directly available but the maintenance costs need to be calculated by a formula taking into account salaries, spare parts and tools.

The cost breakdown structure is system specific. General guidelines for typical cost categories of manufactured systems can be defined (see, e.g., [6]), but, at the least, the lowest level cost elements need to be defined on a case-by-case basis. Table 1 shows an example of the cost breakdown structure for the power supply system, which was the system considered in this case. This cost structure is defined based on practical needs, and it was constructed in close co-operation with researchers and the case company. The practical requirements set for this case have strongly influenced, for example, the cost structure at the beginning of the life phase. There was no point in creating a detailed cost structure for the design and manufacturing phases because the acquisition and installation prices were readily available.

4.2 Estimation of Costs

Several qualitative or quantitative methods can be used to estimate the life cycle costs. Niazi et al. [7] have classified and presented techniques for product cost estimation, but the same methods can also be applied to the life cycle cost estimation. Niazi et al. [7] classified qualitative techniques for intuitive and analogical techniques and quantitative techniques for parametric and analytical techniques. In this case, the total life cycle cost was estimated by the analytical breakdown approach, the execution of which is described in the previous chapter, and which is according to the dependability management standard. In the breakdown approach, the total cost can be worked out from the lower level costs.

Table 1 Cost breakdown structure of the case system

Cost hierarchy level 1	Cost hierarchy level 2	Cost hierarchy level 3	
Life cycle phase	Cost category	Cost element	Cost parameters
Acquisition (Beginning of life)	Cost of acquisition		Components selected in the solution
			Acquisition price/component
	Cost of installation		Installation price/component
Use time (Middle of life)	Cost of maintenance	Corrective maintenance	Number of actions/year/component
			Average cost/action/component
		Preventive maintenance	Number of actions/year/component
			Average cost/action/component
	Cost of energy	Electricity required by the end process	Electrical load of the customer process
		Waste electricity caused by inefficiencies	Efficiency/component
			Electrical load/component
			Use time/component
		Air condition	Amount of waste energy/year
			Multiplier for the need of air condition
		Amount of wind and solar energy	Use time/year/electricity production component
			Load/electricity production component
			Efficiency/production component
		Electricity produced by diesel generators	Electrical load
			Efficiency
			Use hours/year
			Diesel consumption [litres/hour]
			Diesel price/litre

(continued)

Table 1 (continued)

Cost hierarchy level 1	Cost hierarchy level 2	Cost hierarchy level 3	
		Electricity produced by fuel cells	Electrical load
			Efficiency
			Use hours/year
			Used fuel
			Fuel consumption [litres or kg/hour]
			Fuel price/litre or kg
		Fixed costs	Electricity connection fee/year
			Cost of electricity/kWh
	Cost of unavailability	Time of unavailability /year	Number of outages/year
			Average duration/outage
		Unavailability cost /year	Average cost of unavailability/hour
Disposal (End of life)	Disposal of components	Recycling costs	Number of components recycled during system lifetime
			Cost of recycling/component
		Disposal costs	Number of components disposed of during system lifetime
			Cost of disposal/component
	Disposal of the system	Recycling costs	Cost of recycling/system
		Disposal costs	Cost of disposal/system

To estimate the costs at the lower level of the cost hierarchy, both parametric and intuitive techniques were applied. The parametric technique, in which the cost is expressed as a function of its parameters (shown in the last column of Table 1), was used to estimate the cost elements (level 3 in the hierarchy shown in Table 1). These equations can be very simple, e.g. maintenance cost times the number of maintenance actions, or much more complicated, e.g. the calculation of the energy cost. The detailed equations for this case are not shown here because they need to be defined on a case-by-case basis together with definitions of the cost elements.

The numerical values for most of the cost parameters were estimated by an intuitive technique based on expert judgements. Some of the parameter values were available as constants from the price catalogues or the results of reliability tests carried out by the manufacturers.

4.3 Sensitivity Analysis

Life cycle cost calculations are typically used to support decision-making, which means that calculations are performed before the costs are realized and the calculations are based on estimates of future values, which are inherently uncertain. The robustness of the LCC calculation results for the change in cost parameters can be evaluated by a sensitivity analysis. The simplest way to conduct a sensitivity analysis is to change the variable value and then re-calculating the results, i.e. a what-if analysis. In the developed LCC tool, the sensitivity analysis is performed with a Monte Carlo simulation. The result of a sensitivity analysis is the variation in the expected life cycle costs.

4.4 Results of the LCC Calculation

The purpose of the LCC calculation is generally to support decision-making, i.e. to answer the questions a decision-maker has concerning the decision situation. Thus, the result indicators need to be defined according the needs of the decision-maker.

In the current case, the decision situation concerns the purchasing of a power supply system. The technical requirements of a power supply system set by a customer can typically be fulfilled by a number of different solutions when the cost is an important criterion for selection. In this case, it was assumed that the decision-maker, who is a customer of the case company, would ask the following questions:

1. How much would the ownership of a power supply solution or solutions cost?
2. How soon, if at all, would the more expensive investment costs be paid back by the lower annual cost?
3. How would the life cycle costs be divided into different cost factors?
4. What would the variation be of the expected life cycle cost or cost factors?

To answer the first two of the above questions, the total life cycle cost; net present value (NPV), i.e. discounted life cycle cost; annual cost; and free and discounted cumulative annual cost were calculated as result indicators. Question three was answered by presenting the cost elements of the life cycle and annual costs. The last question can be answered after a sensitivity analysis that yields the variation in the expected costs.

4.5 LCC Tool

In principle, the LCC calculations are quite simple and do not necessarily need special calculation tools. When the same kind of calculations need to be repeated it is sensible to create a tool to expedite calculations, which allows calculations to be performed together with customers and even to play a bit and test unusual solutions.

In this case, a prototype tool was developed that enables LCC calculations for different kinds of power supply systems. MS Excel 2010 was selected as the development tool because MS Excel and Visual Basic for Excel include features that provide good support for this kind of prototype development. Another important reason for the selection was that MS Excel is a widely used tool and can thus be used without new IT investments, and potential users are familiar with it.

The developed LCC tool includes cost breakdown structure, data input forms, calculation of result indicators, sensitivity analysis, presentation of results by numbers and graphs. Cost breakdown structure, which was shown in the Table 1, is implemented in the tool as fixed e.g. the user cannot change the cost parameters in the tool. The fixed cost structure could be used here because it is possible to define the relevant costs elements in detailed for a known system. For a general case, this would not have been possible. To make the tool more user-friendly and easier to test in real use cases, an interface with user forms was created for data input. All the calculations are conducted in Excel worksheets and the necessary formulae were implemented.

To support risk-based decision-making, a sensitivity analysis part was implemented in the LCC tool. In this case, the sensitivity analysis was conducted by a Monte Carlo simulation, based on the statistical distribution of the cost parameters. The LCC tool allows a Normal or Weibull distribution to be selected for all variables except the count variables for which only the Poisson distribution is available in this tool. Practitioners who are not typically familiar with the statistical distributions may find it difficult to define the distribution parameters. Thus, a graphical tool was developed (Fig. 2) with which the user can see the shape of the defined distribution and assess if the distribution describes his/her view of the variation.

In the LCC tool, the results of the calculation are shown by tables and graphs (see Fig. 3) in user forms. In addition, the tool allows a report to be created in pdf format that can be given to the customer.

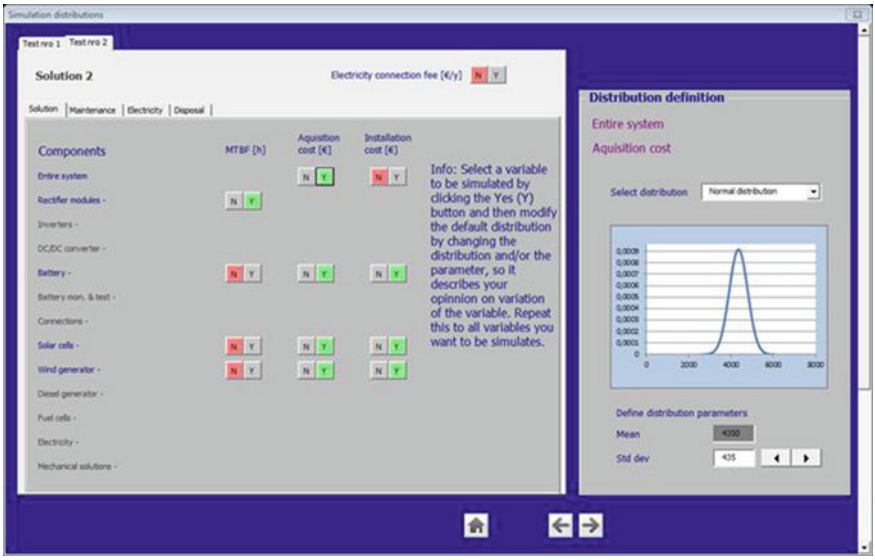


Fig. 2 Example of the definition of a statistical distribution for a sensitivity analysis

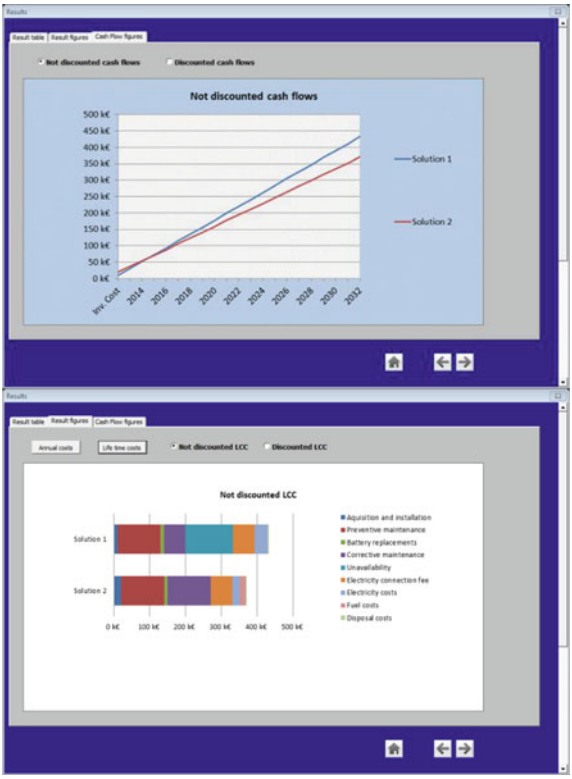


Fig. 3 Example of the result figure types implemented in the LCC tool

5 Use of the LCC Tool

The previous chapter describes how the LCC calculation has been established in this case and the tool developed to support calculation in practice. This chapter presents how the LCC tool can be used.

The use of the LCC tool can be described as a process with five steps:

1. Define the possible solutions that meet the customer's technical requirements and are options to be analysed.
2. Populate the LCC tool with input data, i.e. give numerical values to the relevant cost parameter for current case.
3. Calculate the results. This is done automatically by the LCC tool.
4. Assess the results and compare the options using result indicators from the LCC calculation.
5. Make the decision for the current case based on economic criteria. Other possible criteria from elsewhere can also be used to support the decision.

Life cycle cost calculations can be utilised internally in the company or externally with customers. Figure 4 shows the life cycle phases in which the LCC tool is used internally and externally. The figure outlines the innovation process of a company as part of the life cycle for B-to-B products that are typically delivered by establishing a delivery project with the customer. A company that follows this model develops a product portfolio (consisting of, e.g. services, physical products and product elements that can be configured in different ways to form a final product that meets the customer needs) at its own expense and then sets up projects with clients to sell and deliver products that are based on the elements of its product portfolio and configured and designed to meet the needs of the particular customer.

The life cycle cost calculation can be used in negotiations with potential customers to provide more detailed cost information than just the acquisition price for their decision-making. In this case, the LCC calculations were originally meant to be utilised in the delivery project negotiations with potential customers to serve the case company's need to explain the higher purchasing price with lower life cycle costs and more sustainable solutions. The tool that was developed provides a reasonably

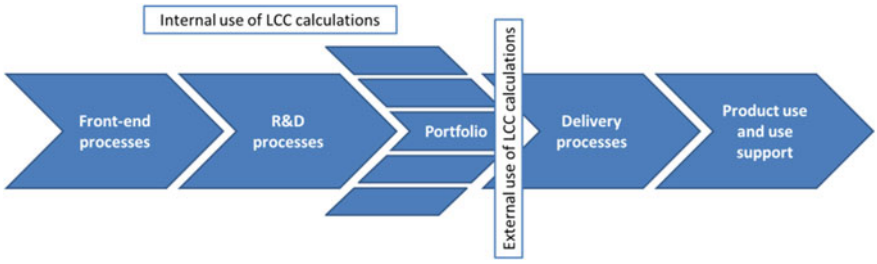


Fig. 4 Uses of the LCC tool within a company with respect to the innovation process/product life cycle (modified from [8])

quick and easy way to review different solutions and it can bring new solutions that differ from the customer's first ideas about the solution into the negotiations.

During the LCC tool development and testing, the case company used the tool to analyse elements of its product portfolio. In these tests, the case company realised that this kind of calculation can elicit ideas to improve products from the life cycle perspective. This internal use of the LCC tool can reveal products that are not good enough from a life cycle point of view and should be replaced with products that lead to better overall results.

6 Conclusions

According to the first preliminary test cases, the customers were very interested in the possibility of having more detailed life cycle cost information. The tool that was developed was seen as a promising way to support decision-making based on life cycle costs. The tool itself was seen as quite easy to use, but the calculations require input data from both the customer and product supplier, and the tool should therefore be used by the customer and product supplier together.

In the preliminary test cases, critical views on the use of this kind of calculation were also raised. The main concern was that all the component suppliers might not be able to provide comparable and reliable input data for LCC calculation if the goal is to compare solutions from several providers. Public purchases, in particular, have strict requirements for transparent competition criteria and all the providers have to be treated equally. It may then be risky to use data that cannot be validated (e.g. future maintenance costs) before decision-making. A sensitivity analysis can be carried out to answer this concern.

The uncertainty concerning future data is present in the decision-making, but it cannot always be used as an excuse for making decisions based on realised and easily validated data, e.g. the acquisition price. In this kind of decision-making it has to be understood that the LCC calculation cannot provide the exact costs but it does provide the estimated magnitude of the total costs and relevant cost categories. This result supports different solutions being put in order and it is a better basis for decision-making than the pure acquisition price.

References

1. Hardin G (1968) The Tragedy of the Commons. Sci New Series 162(3859):1243–1248
2. Gluch P, Baumann H (2004) The life cycle costing (LCC) approach: a conceptual discussion of its usefulness for environmental decision making. Build Environ 39:571–580
3. Götze U, Northcott D, Schuster P (2008) Investment appraisals methods and models. Springer, Berlin
4. Valkokari K, Valkokari P, Rantala T, Palomäki K, Reunanen M & Kunttu S (2013) How to co-create sustainable solutions within manufacturing networks? The XXIV ISPIM Conference—Innovating in global markets: challenges for sustainable growth in Helsinki, Finland on 16–19 June 2013

5. IEC 60300-3-3. (2004) Dependability management—Part 3-3: Application guide—Life cycle costing (2nd edn)
6. Asiedu Y, Gu P (1998) Product life cycle cost analysis: state of the art review. Int J Prod Res 36 (4):883–908
7. Niazi A, Dai JS, Balabani S & Seneviratne L (2006) Product cost estimation: Technique classification and methodology review. Trans ASME J Manufac Sci Eng 128(2):563–575. Publisher: ASME, USA
8. Ahonen T, Reunanen M, Kunttu S, Hanski J, Välisalo T (2011). Customer needs and knowledge in product-service systems development. Comadem 2011 24th International Congress on Condition Monitoring and Diagnostics Engineering Management – Advances in Industrial Asset Integrity Management. Stavanger, NO, 30 May-1June 2011. Det Norske Veritas University of Stavanger; Comadem International Stavanger, NO

Managing Modern Sociotechnical Systems: New Perspectives on Human-Organization—Technological Integration in Complex and Dynamic Environments

Haftay H. Abraha and Jayantha P. Liyanage

Abstract Modern sociotechnical systems (SSs) are becoming increasingly advanced, complex, boundary-less, and technology-dominant systems that have major economic, societal and environmental implications. Digital technologies are enabling us to develop systems with various levels of complexities and interconnections involving different elements. This creates new ways of organizing work, new work processes, for instance: creating closer cooperation across organizational and geographical borders and this trend is likely to increase. Complexities are associated not only with the large scale hardware and software infrastructures, but also with the even more complex issues involved in human and organizational behaviours and characteristics. This implies that there are many hidden risks under the modern systems development and deployment process, and subsequently the potential for serious events are considerable. A major area for study in this context is the establishment of a seamless connection between the characteristics of the individual components (at micro-level) and the macro-behaviour of the complex SSs. Given the complexity of the systems involved, use of classical/traditional approaches (e.g. linear relations of causality) alone to understand the behaviour and performance of these systems are quite challenging, if not extremely limited in use. We need new perspectives to understand the behaviours and interactions in wider context, so that the new perspectives can capture the complex issues that influence Human-Organization-Technological (HOT) conditions within such systems, can emerge. This paper elaborates on an approach that can provide the basis for micro-macro integration to reduce vulnerabilities based on a better awareness (i.e. system thinking) taking into account the dynamic and complex context from a new perspective.

Keywords Complex systems · Human error · Human-Organization-Technology · Unwanted events · Risk · Safety incidents · Sociotechnical system

H.H. Abraha (✉) · J.P. Liyanage
University of Stavanger, Stavanger 4036, Norway
e-mail: haftay.h.abraha@uis.no

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_94

1 Introduction

The stability, performance, and the survival of sociotechnical systems (SSs), as well as their ability to tolerate environmental disturbances, are dependent upon the nature, formation and interaction of the human, organizational, and technological subsystems. Modern SSs are becoming increasingly advanced, complex, boundaryless, and technology-dominant systems that have major economic, societal and environmental implications. Digital technologies are enabling us to develop systems with various levels of complexities and interconnections involving different elements. Complexities are associated not only with the large scale hardware and software infrastructures, but also with the even more complex issues involved in human [1] and organizational behaviours and characteristics. Hence there is a need to explore new ways of thinking to manage modern sociotechnical systems in faces of those new scenarios: system thinking as a complement to the traditional risk and safety analysis.

Given the complexity of the systems involved, use of classical/traditional approaches alone to understand the behaviour and performance of these systems are quite challenging, if not extremely limited in use [2]. Major accidents keep occurring that seem preventable and that have similar systemic causes. The following paragraph, for instance, is quoted from the Deepwater Horizon disaster investigation [3] to show the closely replication of many different disasters.

> In many ways, this disaster (Macondo well blowout (2010)) closely replicates other major disasters that have been experienced by the offshore oil and gas industry. Eight months before the Macondo well blowout, the blowout of the Montara well offshore Australia in the Timor Sea developed in almost the same way-with very similar downstream effects…. Piper Alpha (North Sea) platform explosions and fires (1988)…followed roadmaps to disaster that are very similar to that developed during and after the Macondo well blowout. This disaster [Macondo well blowout] also has eerie similarities to the BP Texas City refinery disaster. These similarities include: (a) multiple system operator malfunctions during a critical period in operations,…., (c) neglected maintenance ,…., (e) inappropriate assessment and management of operations risks, (f) multiple operations conducted at critical times with unanticipated interactions, (g) inadequate communications between members of the operations groups, (h) unawareness of risks, (i) diversion of attention at critical times, (j) a culture with incentives that provided increases in productivity without commensurate increases in protection(safety), (k) inappropriate cost and corner cutting, (l) …., and (m) improper management of change.

Part of the explanation for such replication of different disasters is that the current hazard analysis tools are not designed to analyze dynamic complexity of major incidents, which arise from the interaction between actors (social and technical) and the temporal and spatial gaps between actions and consequences. This is because most traditional causal analysis tools model events and causal factors linearly [4]. Besides, these traditional tools focus on events proximal to the loss.

Rasmussen [5] identified six levels in sociotechnical systems: (a) the government level-6, (b) the regulators and industry association level-5, (c) the management level-4, (d) the company level-3, (e) the staff level-2, and (f) the operation level-1.

Each of these levels represents possible sources of "root causes". The aftermath of most hazardous, large-scale technological systems' accidents have serious and long-lasting economic, safety, health and environmental consequences. Hence, it is reasonable to analyze causes at all these levels to be identified so as to prevent recurrences effectively.

However, the analysis of systemic issues, especially those at company, regulators and industry associations, and government levels, (i.e., macro issues) are complex and dynamic. The complexity at these levels arises because causal factors are inter-related and decisions of actors and the corresponding effects are usually separated in time. Unfortunately, most causal analysis tools, such as those evaluated by [2, 6, 7], view cause and effect linearly and are not designed to model changes in the modern system across time. In other terms they are not designed to analyze the dynamic complexity of the emerging modern systems.

It is proposed in this paper that the traditional/classical causal analysis tools can be used to analyze the incident sequence and causal factors that are more immediate to the incident (micro issues). Key causal factors (macro issues) can then be further analyzed using tools designed to model dynamic complexity. The reason, as mentioned earlier, is that most major system accidents do not result simply from a unique set of proximal, physical events but from drift of the whole SS to a state of heightened risk over time as safeguards and controls are relaxed due to conflicting goals and tradeoffs. The challenge in preventing accidents, according to [8], is to establish safeguards and metrics to prevent and detect migration towards a state of unacceptable risk before accident occurs.

Some major thoughts (or rather motivational factors) for new perspectives would in this context be driven by;

- Safety remains the major concern regarding the design and use of modern and complex SSs in the context of rapidly changing technology.
- The increasing complexity in industrial SSs not only created by the latest developments in digital technologies, but also by other change phenomena related to organizational and human elements [1]
- Dynamic and complex environment (e.g. economic pressures, stringent human-safety-environment(HSE) regulations, etc.) has influences on safety of modern industries
- The decision settings of different stakeholders of complex systems have taken different turns involving multiple approaches and notably conflicting criteria

In light of this, it is important to devote some effort to examining our foundations before proposing some incremental improvement in what we do today (refer also [2, 6] for additional critical reviews on foundational safety issues). Re-examining some underlying assumptions and paradigms in safety is invaluable to identify any potential disconnects with the world as it exists today. The assumptions questioned in this paper involve: (a) definition of safety, (b) accident causal models, and (c) understanding on human and organizational error. Subsequently, alternatives based on system thinking are then proposed.

2 Assumptions Questioned and Alternative Approaches

Re-examining some underlying assumptions and paradigms in safety is invaluable to identify any potential disconnects with the world as it exists today. The assumptions questioned in this paper involve: (1) definition of safety, (2) accident causal models, and (3) understanding on human and organizational error.

Assumption 1: Safety Is Enhanced by Increasing the Reliability of the Individual System Components

An MIT (Massachusetts Institute of Technology) professor, Leveson [9] argues that safety and reliability are different system properties: a system can be reliable and unsafe or safe and unreliable. This misperception is epitomized by HRO (high reliability organizations) researchers who suggest that '*organizations in which the system components*[1] *operate reliably will be safe*' [10–13].This belief is simply not true. In modern complex systems, accidents may result from interaction among perfectly functioning components.

Leveson explained this situation with an example:

> The loss of the Mars Polar Lander was attributed to noise (spurious signals) generated when the landing legs were deployed during descent. This noise was normal and expected and did not represent a failure in the landing leg system. The on board software interpreted these signals as an indication that landing occurred (which the software engineers were told they would indicate) and shut the engines down prematurely, causing the spacecraft to crash into the Mars surface.

According to Leveson, the landing legs and the software performed perfectly (i.e., neither failed), but the accident occurred because the system designers did not account for all interactions between the leg deployment and the descent-engine control software.

In the past, the design of SSs was more intellectually manageable and the potential interactions among components could be thoroughly planned, understood, anticipated, controlled and guarded against [8]. Modern SSs, however, no longer satisfy these properties and system design errors are increasingly the cause of major accidents, even when all the components have operated reliably, i.e., have not failed.

Hence, *safety is system property*, not a component property like reliability [9]. Determining whether an offshore oil and gas plant is acceptably safe, for instance, is not possible by examining a BOP (blow out preventer) in the plant. Conclusions can be reached about the reliability of the BOP without referring to the context in which the BOP is used, but safety of the BOP can only be determined by the relationship between the BOP and the other plant components and its environment, i.e., in the context of the whole.

In systems theory, complex SSs are modeled as a hierarchy of organizational levels, each level more complex than the one below [5]. The levels are characterized

[1] Note: 'component' in this paper refers to both physical and human components of a SS.

by emergent properties that are irreducible and represent constraints on the degree of freedom of components at the level below. *Safety is an emergent property* and unsafe system behaviour is defined in terms of safety constraints on the behaviour of the system components.

Safety should then be viewed, using systems thinking and systems theory, as a *control problem* (problem of enforcing the safety constraints) rather than a failure or reliability problem [9]. Safety incidents occur when component failures, external disturbances, and/or potentially unsafe interactions among system components are inadequately controlled (managed). In basic systems theory, in order to provide adequate control, the *controller (barrier)* must have an accurate model of the *process* it is controlling (see Fig. 1). For both automated and human controllers, the *process model* (for human controllers, this model is commonly called the mental model) is used to determine what control actions are necessary to keep the system operating effectively.

The process model includes assumptions about how the controlled process operates and about the current state of the controlled process. Safety incidents in complex systems often result from inconsistencies between the model of the process used by the controller (barrier) and the actual process state [3]. For instance: the local BP manager on the Deepwater Horizon disaster [3] thought the cement had properly sealed the annulus (he did not notice the positive pressure test) and ordered the mud to be removed; the operators at Texas City [14] thought the level of liquid in the isomerization unit was below the appropriate threshold.

Usually, these process models of the controlled system become incorrect due to missing or inadequate feedback and communication channels. The effectiveness of the safety control structure is greatly dependent on the accuracy of the information about the actual state of the controlled system each controller has, often in the form of feedback from the controlled process.

In modern SSs major accidents rarely have a single root cause but result from an adaptive feedback function that fails to maintain safety as performance changes

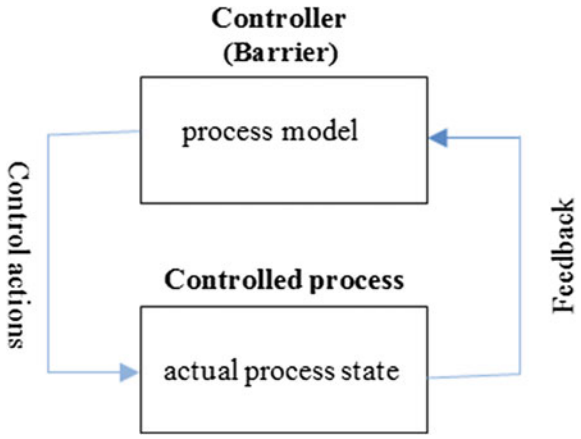


Fig. 1 'Controller-Controlled process' relationship to determine what actions are needed (modified from [9])

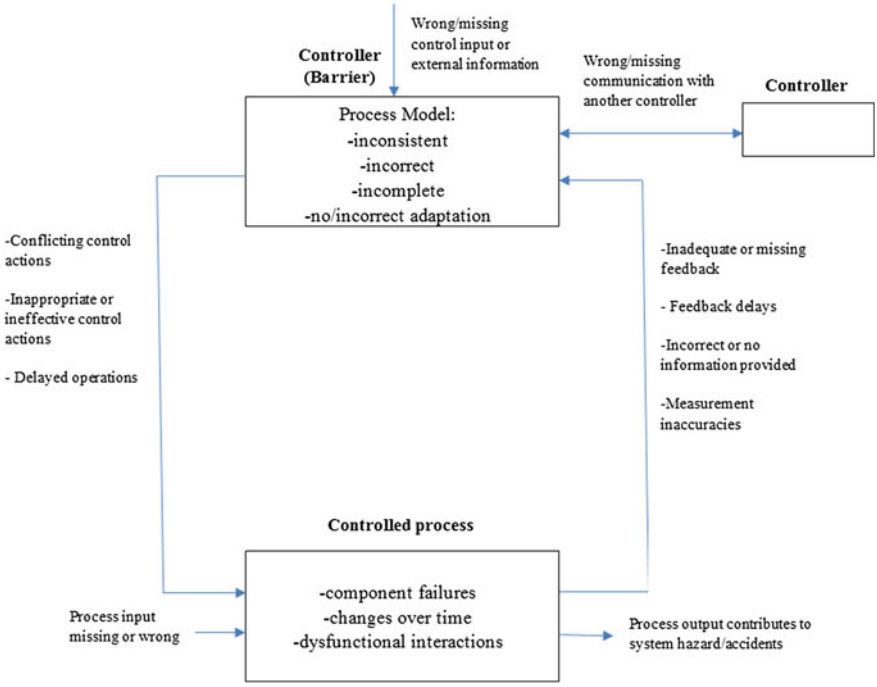


Fig. 2 Some generic factors involved in unsafe control

over time to meet a complex and changing set of goals and values. Figure 2, for instance, shows some of the generic factors involved in unsafe control. Also, as shown in Fig. 2, more than one controller may participate in the safety control structure, with the controllers of the components having individual responsibilities for ensuring that the controlled processes or components are fulfilling their safety responsibilities.

Note that when we say 'control' in our case, it is not only about the controls provided by engineered systems (e.g. interlocks, BOP or various types of barriers and fault tolerance features) and direct management and operational interventions, but also indirectly by policies, procedures, shared values, and other aspects of the organizational culture. Therefore, an accident results not simply from components failure or human error, but from the *inadequate control* of the safety-related constraints on the development, design, construction, management and operation of the entire SS.

Figure 3 below shows the safety control structure existing at the time of the Macondo well system accident. Each component has specific assigned responsibilities for maintaining the safety of the entire SS. For instance, the mud logger is responsible for creating a detailed record of a borehole by examining the contents of the circulating drilling medium, the cementer is responsible for properly sealing off a wellbore, and local management has responsibilities for overseeing that these and other activities are carried out properly and safely. The government oversight

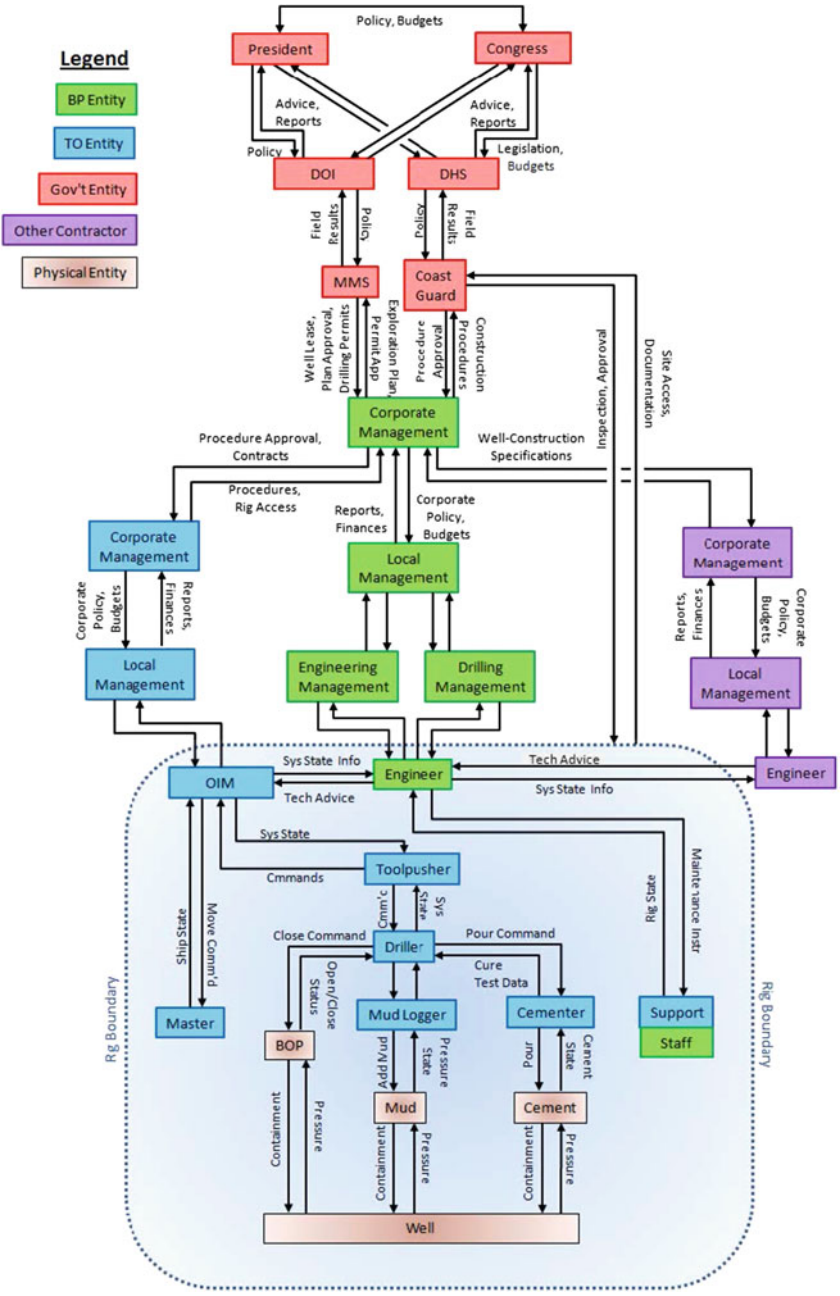


Fig. 3 Safety control structure existing at the time of Macondo blowout (adapted from [8])

agency may be responsible for ensuring that safe practices are being followed and acceptable equipment being used, and so forth and so on.

The main idea that we can draw from this structure is that safety incidents are rarely the result of unsafe behaviour by only one of the components but usually the result of unsafe interactions among and behaviour by all or most of the HOT components in the control structure. A systems thinking approach(such as causal loop diagrams [4], STAMP- Systems-Theoretic Accident Model and Processes [8]) allows capturing the non-linear dynamics of interactions between human, organization and technological (HOT) components of a SS and anticipating the risk-related consequences of change and adaptation over time.

Assumption 2: Accidents Are Caused by Chain of Directly Related Failure Events

This assumption implies that investigating backward from the loss event and identifying directly related predecessor events (usually technical failures or human errors) will identify the "root cause" for the loss. The solution to this approach is then either the "root cause" event is eliminated or an attempt is made to stop the propagation of events by adding barriers between events, by preventing individual failure events in the chain, or by redesigning the system so that multiple failures are required before propagation can occur.

The problem with the chain-of-events model of accident causation is that it oversimplifies causality and the accident process and excludes many of the systemic factors in accidents and indirect or non-linear dynamic interactions among events. It also does not hold for accidents where the cause(s) lies in the interaction among HOT-factors of modern SSs, none of which may have failed.

To hold systemic factors and non-linear dynamic interactions among the HOT factors in modern SSs, the accident causation can be viewed as involving three hierarchal levels, as proposed in Fig. 4. Level 1 is the basic proximate event chain; Level 2 represents the conditions that allowed the events to occur; Level 3 contains the systemic factors that contribute to the conditions and events. The levels are annotated in the figure with the proposed approach to managing modern systems i. e., using systemic thinking approaches (for macro issues) and classical approaches (for micro issues) in a seamless integration.

Classical approaches will help to present facts of the proximal events leading to the loss. The key causal factors (macro factors) can then be further analyzed using systemic thinking approaches. For instance in Macondo blowout: the regulators and the government, each seemed to be doing the "right" thing in view of the pressures that each was facing. However, unintentionally, each actor contributed to the poor safety culture and the worsening of the situation that finally resulted in the blowout. Understanding systemic structure as a whole will help organizations understand the possible negative consequences of their decisions on safety culture and result in the design of more effective safety management strategies. A systems perspective also helps to reduce the tendency to blame a particular group or organization for the incident/accident, thereby increasing the chances of identifying effective proactive measures.

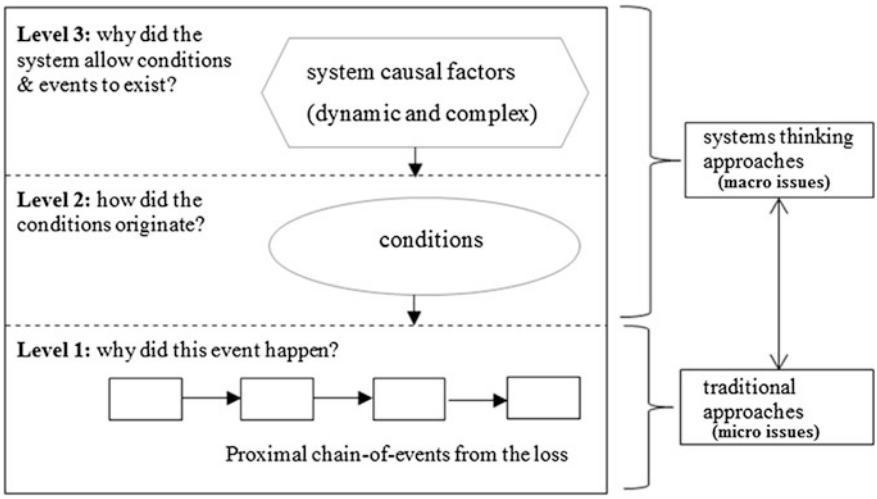


Fig. 4 Macro-micro integration to identify the root cause(s)

The potential advantage of systems thinking is basically to facilitate a more effective way of seeing reality and summarizing dynamically complex situations.

Assumption 3: Most Accidents Are Caused by Human Error

Human behaviour is influenced by the context in which it takes place [15] and hence, changing that context will be more effective in reducing accidents than blaming the human for doing errors. The irony in here is that systems are designed in which human error is unavoidable and then blame the human. Moreover, the present digital, complex and large-scale technological systems (with their dynamic environment) pose additional demands and new requirements on the human operators. Modern SSs require human operators to constantly adapt to new and unforeseen system and environmental demands. Furthermore, there is no clear cut distinction between system design and operation, since the operator will have to match system properties to the changing demands and operational conditions. In other words, according to [16, 17], operators must be able to handle the 'non-design' emergencies, because the system designers could not foresee all possible scenarios of failures and are not able to provide automatic safety devices for every contingency.

Thus, the role of the human operator responsible for such systems has changed from a manual controller to a supervisory controller who is responsible for overseeing one or more computer controllers who perform the routine [15]. In supervisory control systems, the human operator's role is primarily passive, i.e., monitoring of change in the system state. The operator's passive role, however, changes to one of active involvement in cases of unexpected systems events, emergencies, alarm alerts, and/or system failures.

Mental models play a significant role here. The ability to adapt mental models through experience in interacting with the operating system is what makes the

human operator so valuable (see Fig. 5). Designers deal with ideal (or average) systems, and they provide procedures to operators with respect to this ideal. Systems may deviate from the ideal through manufacturing and operation variances or through evolution and changes over time. Operators must deal with the existing system and change their operational procedures using operational experience and experimentation [8]. While procedures may be updated over time, there is usually a time lag in this updating process and operators must deal with the existing system state.

Based on current information, the operators' actual behavior may differ from the prescribed procedures. The irony is that when the deviation brings fortunate results at that particular instant in time, then the operators are considered to be doing their job (and rewarded). However, the operators are often blamed for any unfortunate results, even though their incorrect actions may have been reasonable given the information they had at the time.

Flawed decisions may also result from limitations in the boundaries of the model used, but the boundaries relevant to a particular decision maker may depend on activities of several other decision makers found within the complex modern SS. Safety incidents may then result from the interaction of the potential side effects of the performance of the decision makers during their normal work. It is difficult if not impossible for any individual to judge the safety of their decisions when it is dependent on the decisions made by other people in other departments and organizations [5].

Part of the problem to blaming operators also stems from the linear and deterministic approach to accident investigation where it is usually difficult to find an

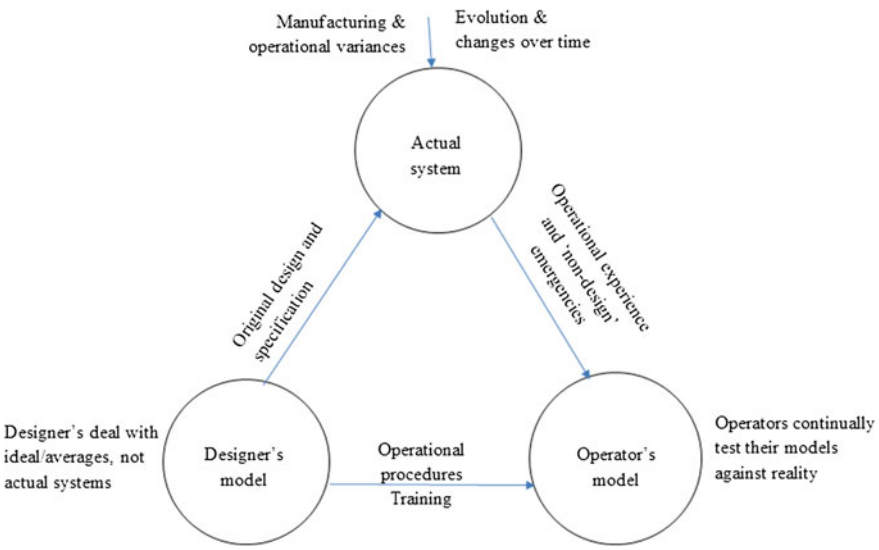


Fig. 5 The role of mental models in operations (modified from [8])

"event" preceding and causal to the operator behaviour [9]. If the problem is the system design, there is no proximal event to explain the error. Even if a technical failure precedes the human action, the tendency is to put the blame on an inadequate response to the failure by an operator. Perrow [11] cites a U.S. Air Force study of aviation accidents that concludes that the designation of a pilot error is a convenient classification for mishaps whose real cause is uncertain, complex, or embarrassing to the organization.

As argued by Reason and others [2, 5, 16–18, etc.], devising more effective accident causality models requires shifting the emphasis in explaining the role of humans in accidents from error (deviations from normative procedures) to focus on the mechanisms and factors that shape human behavior, i.e., the performance-shaping mechanisms and context in which human actions take place and decisions are made. Modeling behavior by decomposing it into decisions and actions (i.e., events) and studying it as a phenomenon isolated from the context in which the behavior takes place is not an effective way to understand behaviour.

3 Analyses and Discussion

Much effort has been done to avoid safety incidents, but they still occur. The problem is that no engineering process is perfect, and every SS and its environment evolve and are subject to change over time. However, our analysis tools are more of static and deterministic (i.e., they model cause and effect linearly and focus on events proximal to the loss).

In general, the causes for safety incidents in modern SSs:

1. May arise in the development and implementation of the system,
2. May reflect management and cultural deficiencies,
3. May arise in operations

3.1 Development and Implementation

- Inadequate safety incident analysis (assumptions about the system hazards or the process used to identify them do not hold)
  - Safety incident analysis is not performed(or is not completed)
  - some safety incidents are not identified or are not handled because they are often assumed to be "sufficiently unlikely"
  - safety incident analysis is incomplete (important causes are omitted)
- Inadequate identification and design of control and mitigation measures for the hazards (e.g., due to inappropriate assumptions about operations)
- Inadequate construction of control and mitigation measures

3.2 Management and Cultural Deficiencies

- The design of the safety control structure is flawed
- The safety control structure does not operate the way it was designed to operate
  - one general cause may be the safety culture, i.e., the goals and values of the organization with respect to safety, degrades over time
  - the behavior of those in the safety control structure may be influenced by competitive, financial or other pressures

3.3 Operations

- Controls that designers assumed would exist during operations are not adequately implemented.
- Changes over time violate the assumptions underlying the design and controls [8]
  - New hazards(arise with changes over time) were not anticipated during design and development or were dismissed as unlikely to occur
  - Physical controls and mitigation measures degrade over time in ways not accounted for in the analysis and design process
  - Components (including humans [1]) behave differently over time (violate assumptions made during design and analysis)
  - The system environment changes over time (violates assumptions made during design and analysis)

To adequately control (manage) safety incidents of modern SSs, we propose systems thinking approaches to work in seamless integration with traditional investigation and causal analysis methods (see Fig. 4). However, the potential advantages of systems thinking approaches to seeing reality and summarizing dynamically complex situations in more effective way depend on the people to adapt a systems-oriented paradigm.

To understand the cause of accidents and to prevent future ones, the system's hierarchical safety control structure (see Fig. 4) must be examined to determine why the controls at each level were inadequate to maintain the constraints on safe behaviour at the level below and why the events occurred. To get a deep enough understanding of the causal factors in an accident such as the Macondo blowout, the reasons for the events and the conditions leading to those events as well as systemic causes need to be identified.

The first step in the safety incident analysis is, hence, to understand the physical proximal factors (micro issues) involved in the loss, including:

- the limitation of the physical system design (e.g., The BOP system was neither designed nor tested for the dynamic conditions that most likely existed at the time that attempts were made to recapture well control),

- the failures and dysfunctional interactions among the physical system components (e.g., the operator at Macondo blow out did not notice the positive pressure test), and,
- environmental factors (e.g., deep water, high temperature, high pressure-HTHP) that interacted with the physical system design

Most classical accident analyses include this information, though they usually omit dysfunctional interactions and look only for component failures. Understanding the physical factors leading to the loss is only the first step in understanding why the accident occurred.

The next step is, understanding how the engineering design practices contributed to the accident and how they could be changed to prevent such an accident in the future. Why was the hazard (e.g., blowout as a result of spills) not adequately controlled in the design? Some controls were installed to prevent this hazard (for example, the BOP and the assignment to see pressure test), but some controls were inadequate or missing.

Many of the reasons underlying poor design and operational practices stem from management and oversight inadequacies due to conflicting requirements and pressures. Identifying the factors lying behind the physical design starts with identifying the safety-related responsibilities assigned to each component in the hierarchical safety control structure along with their safety constraints [8]. Using these safety-related responsibilities, the inadequate control actions for each of the components in the control structure can be identified. In most major accidents, there is inadequate control exhibited throughout the structure, assuming an adequate control structure was designed to begin with (see Fig. 2). But simply finding out how each person or group contributed to the loss is only the start of the process necessary to learn what needs to be changed to prevent future accidents. We must first understand why the "controllers" (see Fig. 1) provided inadequate control. The analysis process must identify the systemic factors in the accident causation, not just the symptoms.

To understand why people behave the way they do, we must examine their mental models and the contextual factors affecting their decision making. All human decision-making is based on the person's mental model of the state and operation of the system being controlled (see Fig. 5). Preventing inadequate control actions in the future requires not only identifying the flaws in the controllers' process models (including those of the management and government components of the hierarchical safety control structure) but also why these flaws existed.

4 Conclusions

Two conclusions may be drawn from the study. One conclusion is related to the critical review on some key assumptions (such as assumptions on: safety vs. reliability; accident causation models; human and organizational error) and the

important of defining safety as: a control problem; system property (i.e., safety deals with systems as a whole, not just components, knowing that many accidents occur because of interactions among HOT subsystems); an emergent property.

The second conclusion is related the proposed approaches (i.e., using systemic thinking approaches as a complement to the classical approaches) to managing risk & safety incidents of modern and dynamic SSs. This approach could remove hindsight bias views and conduct a modern SS towards foresight views.

In the future human, organization and technology (HOT) subsystems will be even more coupled and interdependent, and the boundaries between them will be more blurred. Complex and dynamic interactions, an ever advancing digital technology, trust (both in the technology and human), and common situational awareness among the people at different locations are some of the issues that are more likely to become more important in the future.

Even though traditional causal analysis tools are useful and necessary, they model cause and effect linearly and they are less effective in representing the complex and dynamic interactions between multiple actors and factors across time. It is therefore proposed that systems thinking approaches (methods, tools) should be employed in the analysis of macro-issues in a seamless integration with the traditional tools (which deal with proximal-to-the-loss events/micro issues) so that the systemic structure that contributed to the incident can be more readily understood. The use of systemic thinking approaches could facilitate the early identification of emerging problems in modern industries so as to introduce proactive measures that improve safety and risk management capacity rather than event-level interventions. In addition, it is believed that more research and application of systems thinking concepts will improve the overall effectiveness of safety, health and environment management.

References

1. Liyanage JP (2012) Human-An asset or a liability: The real deal with modern humans in intelligent systems and complex operations, Daejeon, WCEAM2012, accepted
2. Abraha HH, Liyanage JP (2012) Review of theories and accident causation models: Understanding of the human-context dyad towards the use in modern complex systems, Daejeon, WCEAM2012, accepted
3. Deepwater Horizon Study Group, DHSG (2011) [online]. Retrieved from <http://ccrm.berkeley.edu/pdfs_papers/bea_pdfs/dhsgfinalreport-march2011-tag.pdf > [accessed: 03 FEB 2013]
4. Goh YM, Brown H, Spickett J (2010) Applying systems thinking concepts in the analysis of major incidents and safety culture. Saf Sci 48(3):302–309
5. Rasmussen J (1997) Risk management in a dynamic society: a modeling problem. Saf Sci 27 (2):183–213
6. Abraha HH, Liyanage JP (2012) Review of theoretical foundations for risk minimal operations in complex socio-technical systems: The role of human error, Daejeon, WCEAM2012, accepted
7. Sklet S (2004) Comparison of some selected methods for accident investigation. J Hazard Mater 111(1–3):29–37

8. Leveson NG (2013) A systems thinking approach to leading indicators in the petrochemical industry: ESD working paper series. Retrieved from <http://esd.mit.edu/wps/2013/esd-wp-2013-01.pdf> [accessed: 15 May 2013]
9. Leveson N (2011) Risk Management in the oil and gas industry [online]. Retrieved from <http://mitei.mit.edu/news/risk-management-oil-and-gas-industry> [accessed: 10 FEB 2013]
10. LaPorte TR, Consolini PM (1991) Working in practice but not in theory: theoretical challenges of "high-reliability organizations". J Public Admin Res Theory: J-PART 1(1):19-48. Oxford Journal of Public Administration Research and Theory, Inc
11. Perrow C (1999) Y2 K as a normal accident. International Conference on Disaster Management and Medical Relief, June 14-16, Amsterdam
12. Roberts KH (1990) Managing high reliability organizations. Calif Manag Rev 32(4):101–114
13. Weick Karl E, Sutcliffe K, Obstfeld D (1999) Organizing for high reliability. Res Org Behav 21:81–123
14. CBS(2005) Chemical Safety Board (2007). Investigation report: Refinery explosion and fire, BP Texas City, March 23
15. Dekker S (2006) The field guide to understanding human error. Ashgate, Aldershot
16. Reason J (1990) Human error. Cambridge University Press, Cambridge
17. Reason J (1997) Managing the Risks of Organizational Accidents. Ashgate, England
18. Rasmussen J (1986) Information processing and human-machine interaction: an approach to cognitive engineering. North-Holland, New York

Decision Support for Operations and Maintenance of Offshore Wind Parks

Ole-Erik Vestøl Endrerud and Jayantha P. Liyanage

Abstract The world needs cleaner energy, and it needs more to keep up with the growing demand globally. Renewable energy can provide long-term and low emission energy, and wind energy has a large potential. Offshore wind energy capacity in the EU have grown from zero in 1990 to 4 GW in 2011, meeting 0.4 % of electricity demand in the European Union, and the aim is to have 150 GW installed capacity by 2030. However, revenue is an issue in order to make offshore wind energy economically viable in the future, hence, costs must be lowered and at the same time availability must be increased. This paper presents a simulation model developed for research experiments and decision support. It is based on an ongoing project in the North Sea region that investigates configurations of operational infrastructure and work management systems under different governing conditions. The project closely collaborates with one of the largest offshore wind park operators in North Sea, and aims at the use of agent-based and discrete event simulation to experiment with different wind park development scenarios, and to eventually provide decision support for wind park developers and—operators. Despite the use of different modeling techniques in offshore wind sector, the potential benefits of agent-based simulation models in operational planning and work management is still to be explored. The simulation model developed in this paper is based on a multi-method paradigm involving both discrete-event and agent-based modeling. This multi-method approach helps largely in limiting the set of assumptions as well as in managing the drawbacks associated with a specific simulation technique. The paper intends to explain the simulation model developed, discuss the validity of the model and how such models can provide information for decision making in planning and operating offshore wind parks.

Keywords Wind energy · Operation and maintenance · Decision support · Simulation modeling · Agent based modeling · Discrete event modeling

O.-E.V. Endrerud (✉) · J.P. Liyanage
University of Stavanger, Stavanger, Norway
e-mail: ole-erik.v.endrerud@uis.no

J.P. Liyanage
e-mail: j.p.liyanage@uis.no

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_95

1 Introduction

Wind energy has achieved increasing momentum during the last decade due to a rising demand for carbon neutral green energy. While much of the efforts have been on the development and utilization of onshore wind energy sources earlier, for example in Denmark, Germany and USA, the industrial opportunities relating to offshore wind energy have been growing the last few years, and the energy potential in offshore wind is large. In EU alone offshore wind energy capacity have grown from zero in 1990 to 4 GW installed capacity in 2011, meeting 0.4 % of electricity demand in the European Union [1].

Large-scale offshore wind turbine parks (WTP) consist typically of 80–100 wind turbines (WT) composed of more than thousand components each, located from near shore up to 100 km from the closest supply base, and strongly affected by local weather conditions (wind and waves). The typical turbine sizes used in offshore wind parks are in the range 2–3.6 MW, while the turbine size is increasing to 5 MW and successively to an estimated maximum size of 10 MW in future developments. Newly commissioned parks normally include a warranty period (3–5 years), in which the original equipment manufacturer (OEM) has the primary responsibility for operations and maintenance (O&M) in accordance with an O&M contract. After the warranty period operators typically outsource O&M contracts, which are normally partly or fully performance based, dependent on wind park availability and downtime.

However, two of the largest hurdles for the industry at the moment are low margins and high cost of energy (CoE), making subsidizing an important mechanism to stimulate industrial investments (for more on energy support mechanisms in the EU see [2]). An important part of the CoE are operations and maintenance costs (OPEX), which accounts for 20–30 % of CoE for offshore wind parks [3]. Typical OPEX for offshore wind parks in operation in the EU range from 10 to 13 £/MWh [2] and US offshore wind parks range from 17 to 35 £/MWh [4], both, very large figures compared to onshore wind projects in the US, where average OPEX is 7 £/MWh [4]. The high CoE and OPEX for offshore wind compared to onshore wind is much due to the expensive marine logistics and access restrictions in the wind park. With strong winds and high seas, access to a WT and marine operations are very difficult to carry out, and restriction on significant wave height is typically in the order of 1.5–2 m for small personnel transfer vessels (PTV) and large jack-ups. Offshore wind is also competing with other markets for marine logistics, e.g. the oil and gas industry and other offshore wind parks in the North Sea basin, which drive OPEX upwards due to short term vessel availability constraints. Especially jack-ups are a scarce resource.

The choice of maintenance strategy, and configuration of operational infrastructure and work management system will affect availability and OPEX to a great extent, but is dependent on more than the wind park itself. As mentioned earlier local weather affects access to WTs and restricts marine operations in the park. Moreover, outsourcing of O&M activities make the maintenance strategy

dependent on a network of suppliers and service providers, with parts of the operational infrastructure depending on service providers. In addition, when O&M is dependent on an O&M service supplier network work management must be seen in a more integrated approach for efficient and effective coordination and planning of activities.

Different maintenance strategies and configurations of operational infrastructure and work management system have been proposed and investigated. Especially fleet mix and logistic solutions [5–6] and condition based maintenance [7, 8].

Because the offshore wind industry is a fairly small and young industry only limited experience and historical data are available. In order to conduct research on this problem a research method that is not dependent on historical data should be used. Simulation modeling was chosen as research method to look at an offshore wind park in a long-term perspective (10–20 years). A validated simulation model can generate data series, which can be studied to investigate the effect of different configurations of operational infrastructure and work management systems under different maintenance strategies, and their effect on OPEX. The highly complex setting in the offshore wind industry together with the lack of experience, historical data and unsuitability for large scale testing make this an excellent subject for simulation modeling.

Several other simulation models for offshore wind O&M exist. Norwegian offshore wind cost and benefit model (NOWIcob) is a simulation model and decision tool developed by SINTEF for the purpose of simulating O&M and marine logistic support during full lifetime cycles of an offshore wind park, to establish life cycle cost and profit [9, 10]. This is a simulation model that also includes a Markov Chain weather model based on historical data to include weather uncertainty [11]. The University of Strathclyde develops another simulation model similar to the one explained in this paper, which is based on MATLAB and uses Bayesian Belief Networks for decision modeling [12]. Besnard et al. [6] have developed analytical models to evaluate maintenance strategies and operational infrastructure concepts. A thorough review of other O&M and cost simulation models for offshore wind application is presented in [13]. Despite the apparent interest for agent-based simulation in other fields of research such as ecology, economy and business strategy [14] there have been surprisingly little work done in the field of maintenance management. Only Kaegi et al. [15] have analyzed O&M strategies by means of agent-based simulation.

In order to make offshore wind energy economically viable in the future, OPEX and consequently CoE must be lowered by increasing wind park availability through reducing downtime. To achieve this ambitious target new ideas and thinking must be introduced in terms of work processes, maintenance strategies and intelligent technologies. This paper suggests a new simulation model based on a multi method approach to investigate maintenance strategies, operational infrastructure and work management systems in a long-term perspective. Section 2 is a description of the simulation model, Sect. 3 contains the first preliminary results used for verification and validation, as well as investigating operational

infrastructure for a corrective/preventive maintenance strategy, and Sect. 4 describes possible application areas for such a model.

2 Model Description

Simulation modeling is a method to study the real world in a computerized model where controlled experiments can be conducted. In comparison to analytical models a simulation model can capture internal variability and dynamic relations.

A multi method modeling approach is used combining agent based (AB) and discrete event (DE) modeling to capture individual characteristics of turbines, vessels and crew, in addition to decision making by the two latter, while efficiently modeling work processes. An agent-based modeling approach is a bottom-up modeling method that captures the individual characteristics of all individuals in a system and its environment [16]. This modeling approach is—as opposed to discrete event simulations and system dynamics—more capable of capturing intelligent behavior in systems involving active, autonomous and adaptive entities (e.g. organizations, people, machines, stocks), especially with emergent characteristics [17]. On the other hand, DE is excellent at modeling processes, especially work processes in this simulation model, but the entities passing through the process do not exhibit individual characteristics; thus being a top-down modeling method. In this model AB and DE are built into each other. It is important to keep in mind that these modeling paradigms must not be seen as disparate modeling methods, but rather as two outer poles on a continuum where different modeling problems can be located in between.

While several software packages exist for agent-based (AB) and discrete event (DE) simulation modeling (NetLogo, Arena, etc.), Any Logic is used for this project because of its ability to very well integrate AB and DE paradigms in the same model. Any Logic models are based on Java and the full code can easily be debugged and viewed for verification purposes.

This section will present an overview of our O&M simulation model for offshore wind with required input data, the internal model logic and the resulting output data.

2.1 Input Data

Our simulation model provides a framework for testing and optimizing maintenance strategies, conduct parameter-sensitivity analyses, and optimizing work—and decision processes. But it requires a comprehensive set of input parameters and input data. It is also important to state that this model does not include the grid or converter system.

2.1.1 Weather Data

A bi-variate Markov Chain Monte Carlo model is used in this model to generate synthetic weather time-series, for more information see e.g. [11]. The historical time-series used as a basis to generate the preliminary results are data extracted from the FINO database, which are wave and wind data measurements from outside the German coast. Weather data need to be gathered from meteorological measurement stations in the vicinity of the wind park in focus. The simulation model described in this model uses hourly resolution data for wind speed and significant wave height.

2.1.2 Cost and Electricity Price Data

Cost of spare parts, maintenance personnel and vessel chartering are the main cost components going into the OPEX. In addition, lost production due to WT downtime is also included in the model to provide a measure on the cost of downtime. Income is only based on power production. In order to get reliable cost estimates from the simulation model relevant cost figures and electricity price must be gathered. In this model a 10 year average of 90 £/MWh is used, but more favorable would be to have time series of electricity price relevant for the specific market. The cost of different maintenance actions and spare parts in the preliminary simulations is a combination of expert judgment and cost numbers from the NREL WindPACT project [18]. Perfect demand of electricity is also assumed.

2.1.3 Vessel Data

The vessels represented in this model are generic and are not representing any specific vessels on the market. Four vessel types are used: small personnel transfer vessels (e.g. Windcat), medium size supply vessels with crane capability, large jack-up vessels, and helicopter. The only two vessel characteristics used at this moment are significant wave height criteria (H_s) and wind speed criteria (U_w). Table 1 summarizes vessel characteristics. Whether vessels are chartered in for limited periods or on long-term contracts must be identified. A parameter that is important for transit time is vessel speed, which can easily be found in the vessel's datasheet.

Table 1 Vessel characteristics

Vessel	H_s	U_w	v
PTV	1.5 m	N/A	20 knots
Supply vessel	1.5 m	10 m/s	12 knots
Jack-up vessel	1.8 m	10 m/s	11 knots
Helicopter	N/A	11 m/s	135 knots

2.1.4 Wind Turbine and Park Data

Two important WT input data needed to run the simulation model are failure data and the WT power curve (for more on analysis of failure data see e.g. [19, 20]). As will be explained in the next section the failure model of a WT is a non-homogenous Poisson-process with a failure intensity that varies with time. Five failure categories exist with different failure intensities. All failure intensity-values are found through data analysis of historical failure data. At this moment such historical failure data are not easily available for offshore wind parks, but [21] provides a study of WT failure rates, both on component and system level, for onshore turbines. For simulations relating to a specific WTP failure data corresponding to a similar WTP is needed to capture how maintenance tasks are generated; consequently affecting what type of maintenance strategy should be used and operational infrastructure suitable for this wind park.

Another WT and park specific data is the power curve. The power curve determines how much electrical power a WT will produce at different wind speeds, and is dependent on turbine manufacturer, turbine capacity and location in the park. In general there exist a cut-in wind speed ($U_{cut\ in}$) denoting when a WT starts producing electricity from the wind, and a cut-out wind speed ($U_{cut\ out}$) when wind speeds are to high and electricity production stops (see Fig. 1), e.g. by stopping the rotating rotor or ensuring zero lift force by pitching the blades out of the wind for variable speed WTs. The power curve also specifies what the rated power is for a WT, corresponding to the machine plate capacity, and at what wind speed this power maximum is reached. Because power curves are specific for each turbine model and—manufacturer this is an important input data that need to be provided for each WTP.

2.2 Model Logic

2.2.1 Failure Module

WTs and its subassemblies are complex mechanical systems and the number of failures between time 0 and t, $N(t)$, is commonly modeled as a Power Law Process [19], where the non-homogenous Poisson-process is a special case if minimal repairs are assumed [22], expressed mathematically in Eq. (1)

$$P(N(t) = i) = \frac{(\lambda(t) \cdot t)^i}{i!} e^{-\lambda(t) \cdot t}, \quad i = 0, 1, 2, \ldots \tag{1}$$

with failure intensity $\lambda(t)$ at time t (average number of failures per time unit t) represented by a Weibull function

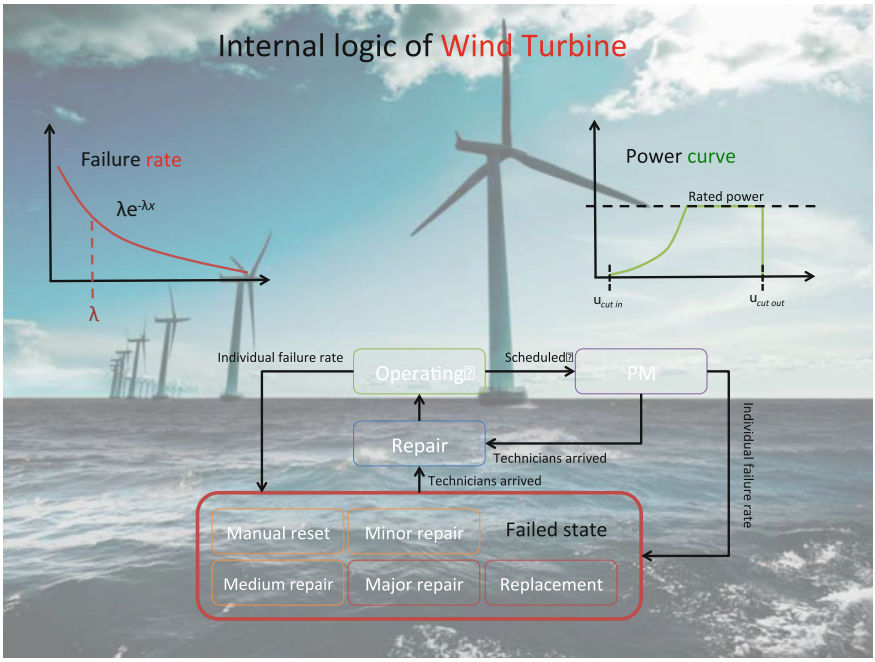


Fig. 1 State chart and individual characteristics of the WT agent

$$\lambda(t) = \lambda\beta t^{\beta-1} \tag{2}$$

where λ is the annual failure intensity when the turbine is new or repaired β is the distribution shape factor (see [22, 23] for more details). If β equals 1.0 the failure intensity becomes constant in time and reduces to an exponential distribution. A $\beta > 1$ results in increasing failure intensity, typical for the late life phase, and $\beta < 1$ result in diminishing failure intensity, typical for the burn in period in the early life phase. For the preliminary simulations a β value of 0.95 is chosen based on expert judgment, resulting in diminishing failure intensity in the first 10 years of operation. The shape and scale parameter can be estimated from failure data relevant to the simulated wind park. Formula (2) can be rewritten as

$$TTF(t) = \frac{TTF}{\beta t^{\beta-1}} \tag{3}$$

where TTF is Time To Failure when a WT is new, which for mechanical and electrical components is exponentially distributed with parameter $\bar{\lambda}$, the average annual failure rate.

Minimal repair is assumed because WTs consist of a great number of components and repairing or replacing a component will most likely take the WT back to the condition prior to the failure. This is a valid assumption according to [24].

Table 2 Average annual failure intensity

Failure category	λ
Manual reset	7.5
Minor repair	3.5
Medium repair	0.275
Major repair	0.04
Replacement	0.08

The chosen failure model formulation is assumed to be binomial, where a turbine is regarded as one item with the two possible states: (1) *working* or (2) *not working*. With reference to Fig. 1a WT is in the *not working* state when failed or when being repaired, and in the *working* state when operating or waiting for periodic maintenance.

Every turbine can go into the not working state by means of five different failure categories: (i) manual reset, (ii) minor repair, (iii) medium repair, (iv) major repair and (v) replacement. All the five failure categories have individual average annual failure intensities, $\bar{\lambda}_i(t)$, and after a turbine is repaired a TTF_i is sampled from a cumulative exponential distribution with parameter $\bar{\lambda}_i(t)$ for each failure category for the repaired WT. TTF_i is also sampled at the initiation of a simulation run for every turbine. This procedure ensures that every WT will act as one individual turbine, i.e. as one agent in the simulation model, and accounts for uncertainty in actual technical condition. The average failure intensities used in the preliminary simulations are given in Table 2.

A WT goes into the failed state when a timer (model time subtracted by time since last repair) exceeds one of the five TTF_i. All failure categories are independent.

An alternative failure modeling approach is a damage accumulation model applying physics of failure (e.g. fracture mechanics) instead of statistical models, see e.g. [7].

2.2.2 Maintenance Planning and Scheduling Module

In addition to a failure module, the simulation model has a maintenance planning and scheduling module that is analogous to a maintenance manager in real life. The maintenance manager is modeled as an agent which: (1) monitor the WTP and interpret alarms, (2) schedule repairs (create work orders), (3) charter vessels and (4) assign work orders (WO) to vessels.

All personnel, maintenance technicians and—managers, work according to a shift plan, which in this simulation model is set to one 12-h shift per day. Moreover, this means that work will only be performed between 7 a.m. and 7 p.m. every day.

Maintenance scheduling is done by "first-come-first-served" principle, meaning there is no prioritization, except from preventive tasks, which is scheduled after repairs. This is done because a repair action is equivalent to a failed WT that do not produce electricity.

2.2.3 Repair Module

Appropriate vessels, for transport and heavy lifting, and maintenance technicians perform repairs and preventive actions. A WO is assigned to a vessel that will mobilize and sign on as many technicians as needed to complete one or several WO's. Mobilization includes vessel transit time from its previous job to the supply base, sign-on of technicians, and preparation of parts and equipment at the supply base. After mobilization is done the captain check if there is an appropriate weather window to complete the maintenance task. The weather window is defined as the needed time to complete the operation for which the limiting criteria (H_s and U_w), OP_{LIM}, multiplied with an uncertainty factor α are fulfilled [25], see (5). The needed weather window is taken as the total time to complete the WO and transit time to and from the turbine, shown in (4),

$$t_{ww} = t_{transit} + t_{access} + t_{diagnose} + t_{repair} \tag{4}$$

$$OP_{WW} = OP_{LIM} \cdot \alpha \tag{5}$$

where OP_{WW} is the weather window criteria. In (5), α is taken as 1.0 as perfect weather forecast is assumed for the preliminary simulations. For α factors in case of weather uncertainty see [25].

Transit time is dependent on vessel speed and WT location, and because every turbine has a spatial location in 2D space transit time is calculated based on travelling in a straight line between the base and turbine. Access—and diagnosis time are deterministic numbers depending on the maintenance task at hand. Repair time is a probabilistic number with a triangular distribution with maximum, minimum and most likely values depending on the maintenance task. Table 3 lists the average repair, diagnose and access parameters used to in the preliminary simulations, while the OP_{LIM} are given in Table 1.

Table 3 Time parameters

Maintenance action	t_{access} (min)	$t_{diagnose}$	t_{repair} (h)
Manual reset	15	0 min	3
Minor repair	15	1 h	3
Medium repair	15	1 h	10
Major repair	15	1 h	30
Replacement	15	1 h	5

2.3 Output Data

The simulation model generates output in accordance with key performance indicators (KPI) used in the offshore wind industry. Total power production (TP) is given in TWh and total income (TI) is given as the product of TP and a fixed, average electricity price (90 £/MWh used for the preliminary results). Operational expenditure (OPEX) is split in three categories: (i) vessel cost, (ii) spare part and repair cost, and iii) labor cost. Vessel cost is the product of day rate (£/day) and total days chartered (including mobilization time). Spare part and repair cost for each failure category is a deterministic number based on expert judgment, and is an extra cost per WO (£/WO). Labor cost is the product of the number of maintenance technicians and—managers and a fixed yearly salary (£/year). OPEX per produced kWh is also automatically output and is calculated as the quotient of OPEX and TP (£/kWh). To be able to look at maintenance efficiency and impact on production, lost production (LP) is calculated as the electrical power (TWh) not produced due to WT downtime. Technical availability is used to measure maintenance quality. Technical availability is defined as the percentage of time a WT is ready to produce electrical power when the wind speed is between cut-in and cut-out speed [23]. In mathematical terms technical availability is

$$A_{tec} = \frac{\tau - \eta}{\tau} \tag{6}$$

where τ is the total time wind speed is between cut-in and cut-out speed during a time period, and η is the total downtime during the same period. Alternatively availability can be expressed by energy terms

$$A_{energy} = \frac{ATP}{TP} \tag{7}$$

where ATP is actual total production in TWh and TP is the total production in TWh when assuming 100 % technical availability between the cut-in and cut-out wind speed. Faulstich et al. [26] investigated the difference between these two measures of availability and found that the difference is marginal.

In addition, an important measure of WTP utilization is the capacity factor, which is defined as the ratio of the actual production over a period of time and the potential production if producing at full nameplate capacity for the same period of time [23]. In mathematical form

$$C = \frac{TP}{\tau \cdot n \cdot P} \tag{8}$$

where n is the number of WTs and P is the nameplate capacity. The last output is vessel utilization, which, is the ratio of the time a vessel is executing a WO (mobilization, transit and repair) and the total simulation time.

3 Results

The simulation model is still in an infant level, and is currently at the verification and validation stage. Historical data are scarce for the offshore wind energy sector, as only a limited number of WTP are in commercial operation and those who are have not been operating for long periods. This fact makes validation difficult. As goodness-of-fit tests are unviable face validity tests, comparison with similar simulation models and parameter-sensitivity analysis is used to check model validity. Face validity tests have been run with participants from industrial collaborators of the project, and the model was accepted as a credible representation of a real wind park. Consequently, this simulation model will be used in the early concept study for another large offshore wind park project in the North Sea, lead by one of the industrial collaborators. A comparison with similar models from SINTEF, MARINTEK, EDF and University of Strathclyde is carried out at the moment, and is not yet completed at the time this paper is written.

The results of the parameter sensitivity analysis are shown in Table 4. The maintenance strategy used in these simulations is a corrective/preventive maintenance strategy and the parameters varied are: (i) number of personnel, (ii) number of vessels, (iii) failure rate and (iv) weather sensitivity. To assure a satisfying level of precision in the simulation results 50 simulation runs were conducted for each case, and the following convergence criterion were used

$$\frac{\sigma}{\mu \cdot \sqrt{N}} < 2\,\% \tag{9}$$

where σ is the standard deviation of the N runs, μ is the average of the N runs. For all cases in Table 4 the convergence criteria were met.

4 Discussion

The parameter sensitivity analysis shows that the model act as expected in terms of directionality. Because of very rough weather technical availability is low for the base case. Availability gets worse for fewer technicians and vessels, and higher failure intensities—as expected. The explanation is that fewer technicians and vessels means fewer WO's can be completed per time unit, hence, downtime increases. And increasing failure rates will generate more WTs shutting down with the same number of technicians and vessels, also increasing downtime.

Increasing the number of technicians does not have a large impact compared to the base case. This indicates that the vessel availability and weather restrictions are limiting the ability to utilize the added number of technicians. This dynamic behavior is something that is captured in simulation models, but is invisible in analytic models due to the latter's lack of dynamic capability. OPEX is not very different for 10 and 30 technicians; this can be explained by the coordination of

Table 4 Results from parameter sensitivity analysis (corrective/preventive maintenance strategy) after 50 simulation runs for seven different cases

Measure	Base case	10 pers.	30 pers.	1 PTV	No OP_{LIM}	λ −50 %	λ + 200 %
A_{tec}	80.80 %	71.40 %	80.80 %	54.20 %	93.70 %	91.00 %	53.30 %
A_{energy}	80.50 %	71.30 %	80.50 %	54.10 %	93.70 %	90.90 %	53.10 %
OPEX	£169.14 m	£155.76 m	£178.94 m	£138.15 m	£136.70 m	£121.51 m	£235.31 m
OPEX (£/MWh)	0.0193	0.0201	0.0205	0.0235	0.0134	0.0123	0.0408
TP (kWh)	8.74E + 09	7.73E + 09	8.74E + 09	5.87E + 09	1.02E + 10	9.86E + 09	5.76E + 09
LP	£190.01 m	£280.81 m	£190.09 m	£448.28 m	£61.60 m	£89.00 m	£458.14 m
LP (kWh)	2.11E + 09	3.12E + 09	2.11E + 09	4.98E + 09	6.84E + 08	9.89E + 08	5.09E + 09
Vessel cost	£97.46 m	£98.09 m	£98.77 m	£83.30 m	£64.66 m	£67.05 m	£141.61 m
Spares cost	£54.65 m	£48.65 m	£55.12 m	£37.82 m	£55.01 m	£37.43 m	£76.67 m
Labor cost	£17.03 m	£9.02 m	£25.05 m	£17.03 m	£17.03 m	£17.03 m	£17.03 m
Util. PTV	57.5 %	57.9 %	57.5 %	42.7 %	100.0 %	58.0 %	56.5 %
Util. FSV	6.7 %	6.6 %	6.6 %	6.4 %	7.2 %	6.1 %	7.5 %
Util. Jack-up	13.0 %	12.9 %	13.3 %	12.5 %	13.8 %	9.3 %	18.5 %
# of failures	8,080	7,592	8,081	5,888	8,752	4,442	12,560
C	0.416	0.368	0.416	0.279	0.483	0.469	0.274

tasks. With more technicians more WOs will be taken care of in parallel, hence, vessel cost is not increased. However, the slight increase in OPEX for 30 technicians is leveled by the decreased downtime.

Decreasing failure rates is very efficient for increasing availability, meaning designing more reliable WTs is a very good measure for lowering downtime, but must be seen in combination with an increased investment cost. The result of decreasing failure rates is a lower OPEX, actually lower than actual reported OPEX for offshore wind today in the EU (10–13 £/MWh).

Weather criteria are the second most critical parameters, having the potential to increase availability to around 94 %. Moreover, the PTV is the most important vessel and reducing the number of PTVs to one decreases availability drastically and results in very high OPEX. Consequently, increasing the weather criteria of PTVs has a very large potential for increased availability and lower OPEX.

Looking at the number of failures generated these numbers look odd at first, but make sense. With one PTV the number of failures are very low compared to the expected number of failures, but this is because a turbine will move into the "failed"-state and will remain there until repaired, thus, unable to generate more failures before repaired. Therefore, few PTVs mean few repaired WTs and thus few failures generated. Consequently, all these results indicate that the model logic is working.

5 Application Areas

The above section provides evidence that simulation modeling, and this model in specific, represents the real world quite well. In other words, this can be a decision support tool in different life cycle phases of a WTP. This simulation model can provide basis for decisions on: (1) maintenance strategies to use by implementing maintenance scheduling and work processes for testing how it will affect cost and production; (2) which supply base to use or invest in can also be investigated by means of simulation modeling by implementing restrictions and location of specific supply bases; (3) also effects of synergies between closely located WTP, for example by coordinated maintenance activities and sharing vessel capacity; (4) one can also implement contract restrictions and prices to make decisions on contract strategy for maintenance services and vessel chartering; (5) support investment decisions, especially investments in operational infrastructure (vessels, supply bases, personnel, etc.).

6 Conclusion

This paper has presented a simulation model based on a multi-method modeling approach using a combination of agent-based and discrete event modeling. The simulation model is capable of simulating the O&M and logistic support system

during the O&M life cycle of a large-scale WTP. Three modules make up the model logic: (1) failure module, (2) maintenance planning and scheduling module and (3) repair module. Early verification and validation provide evidence for at least first order of approximation, and the preliminary results indicate that directionality and cost numbers are valid. A very important problem is the lack of historical data to validate the model. Furthermore, after investigating a corrective maintenance strategy two parameters emerged as critical: weather criteria for PTVs and WT reliability. Future investments in technology development should therefore be focused around increasing weather criteria (wave height and wind speed) for small PTVs, in addition to increasing the reliability of WTs. Furthermore, a simulation model like this one could also provide new insights into the dynamics of O&M of offshore wind parks, as experience with such industrial assets is very limited. Lastly, this simulation model can be used as a decision support for practitioners in wind park planning and management.

Acknowledgment This work is a part of the Norwegian Centre for Offshore Wind Energy (NORCOWE).

References

1. European Wind Energy Association (2011) Wind in our sails—the coming of Europe's offshore wind energy industry. European Wind Energy Association
2. European Wind Energy Association (2009). The economics of wind energy—a report by the European wind energy association. European Wind Energy Association
3. Engels W, Obdam T, Savenije F (2009) Current developments in wind. Energy research Centre of the Netherlands (ECN), Research report ECN-E–09-96
4. International Renewable Energy Agency (2012) Renewable energy technologies: cost analysis series—Volume 1: power sector. International Renewable Energy Agency, Working Paper
5. Scheu M, Matha D, Hofmann M, Muskulus M (2012) Maintenance strategies for large offshore wind farms. Energy Procedia 24:281–288
6. Besnard F, Fischer K, Tjernberg LB (2013) A model for the optimization of the maintenance support organization for offshore wind farms. IEEE Trans Sustain Energy 4(2):443–450
7. Nielsen JJ, Sørensen JD (2011) On risk-based operation and maintenance of offshore wind turbine components. Reliab Eng Syst Saf 96(1):218–229
8. Tian Z, Jin T, Wu B, Ding F (2011) Condition based maintenance optimization for wind power generation systems under continuous monitoring. Renew Energy 36(5):1502–1509
9. Hofmann M, Heggset J, Nonås LM, Halvorsen-Weare EE (2011) A concept for cost and benefit analysis of offshore wind farms with focus on operation and maintenance. In: Proceedings of the 24th international congress on condition monitoring and diagnostics engineering management, Stavanger, Norway, p 8
10. Hofmann M, Sperstad IB (2013) NOWIcob—a tool for reducing the maintenance cost of offshore wind farms. In: DeepWind 2013. Trondheim, Norway
11. Hagen B, Simonsen I, Hofmann M, Muskulus M (2013) A multivariate markov weather model for O&M simulation of offshore wind parks. In: DeepWind 2013. Trondheim, Norway
12. Dinwoodie I, McMillan D, Revie M, Dalgic Y, Lazakis I (2013) Development of a combined operational and strategic decision support model for offshore wind. In: DeepWind 2013. Trondheim, Norway

13. Hofmann M (2011) A review of decision support models for offshore wind farms with an emphasis on operation and maintenance strategies. Wind Eng 35(1):1–16
14. Bonabeu E (2002) Predicting the Unpredictable. Harv Bus Rev
15. Kaegi M, Mock R, Kröger W (2009) Analyzing maintenance strategies by agent-based simulations: a feasibility study. Reliab Eng Syst Saf 94(9):1416–1421
16. Railsback SF, Grimm V (2012) Agent-based and individual-based modeling: a practical introduction. Princeton University Press, Princeton
17. Borshchev, A Filipov A (2004) From system dynamics and discrete event to practical agent based modeling: reasons, techniques, tools. In: The 22nd international conference of the system dynamics society, Oxford, England
18. Malcolm DJ, Hansen AC (2006) WindPACT turbine rotor design study. National Renewable Energy Laboratory, Subcontract report NREL/SR-500-32495
19. Tavner PJ, Xiang J, Spinato F (2007) Reliability analysis for wind turbines. Wind Energy 10 (1):1–18
20. Guo H, Watson S, Tavner P, Xiang J (2009) Reliability analysis for wind turbines with incomplete failure data collected from after the date of initial installation. Reliab Eng Syst Saf 94(6):1057–1063
21. Faulstich S, Hahn B, Tavner PJ (2011) Wind turbine downtime and its importance for offshore deployment. Wind Energy 14(3):327–337
22. Aven T (1992) Reliability and risk analysis. Elsevier Applied Science, London
23. Tavner P (2012) Institution of engineering and technology, Offshore wind turbines reliability, availability and maintenance. Institution of Engineering and Technology, London
24. Baker RD (1996) Some new tests of the power law process. Technometrics 38(3):256
25. Det Norske Veritas (2011) DNV-OS-H101: marine operations, General
26. Faulstich S, Lyding P, Tavner P (2011) Effects of wind speed on wind turbine availability. In: EWEA 2011. Brussels, Belgium
27. Maples B, Saur G, Hand M, van de Pieterman R, Obdam T (2013) Installation, operation, and maintenance strategies to reduce the cost of offshore wind energy. National Renewable Energy Laboratory, Research report TP-5000-57403

Dealing with Uncertainty in the Asset Replacement Decision

Ype Wijnia

Abstract To prevent dangerous situations from gas leaks, operators of the distribution grids in the Netherlands are required to inspect all pipelines for leakages once every 5 years. Leaks are generally fixed when encountered. However, given that the distribution grid has been in use for many years, it may be wiser to replace leaking sections of the grid. The right choice depends on the costs of replacement versus the expected costs (including the monetized risks) of future repairs. The number of future repairs to expect is uncertain, especially given that the asset is ageing and the failure rate may be rising. In the paper, decision making on a representative case is explored. In the basic approach (using past performance for future failures) repairing is marginally better than replacing. However, assuming an increasing failure rate because of ageing will tip the balance in favour of replacement at some (uncertain) moment in future. The decision problem is thus transformed from a choice between alternatives into timing one specific alternative. To find the optimal moment given the uncertain development of the failure rate both sensitivity analysis and real options analysis are applied. Different assumptions and different decision methods result in different optima. However, from a total cost of ownership perspective the differences are relatively small and they hardly justify the analysis effort: any choice will be acceptable. The paper thus reaches the (surprising) conclusion that it may be better to flip a coin than to try to find the best solution.

Keywords Asset replacement · Asset strategy · Risk management · Uncertainty · Real options

Y. Wijnia (✉)
Asset Resolutions B.V, P.O. Box 30113, 8003 CC Zwolle, The Netherlands
e-mail: ype.wijnia@assetresolutions.nl; y.c.wijnia@tudelft.nl

Y. Wijnia
Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, The Netherlands

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_96

1 Introduction

In the Netherlands, virtually everybody (some 7 million customers) has been connected to the gas distribution infrastructure. This large scale roll out of the system followed on the discovery of large amount of natural gas in Groningen (the Slochteren Field) in 1953, when the decision was made to sell it to consumers instead of industry [1]. However, even though methane is a relatively safe gas (not toxic, not corrosive, not vey flammable), it can explode at the right concentration in a confined space. As natural gas has no smell (instead of the previously used manufactured gas) the build-up of dangerous concentrations can happen unnoticed, with the potential for disaster. A tragic example is the explosion of 1937 in New London, Texas, where an undetected natural gas leak lead to the destruction of the Consolidated High School, killing 294 people [2]. Since then, it is good practice (which later became a legal requirement) to add an odorizing agent to natural gas. A second measure against the dangerous build-up of gas is running a periodic gas leak detection program. The prescribed interval in the Netherlands is 5 years. Detected leaks are classified dependent on the size of leak, which is determined based on the measured concentration in the ground. For large leaks immediate action is required, but for smaller leaks a more delayed response is accepted. The standard response for a leak is to repair is, either by placing a fix over the leak or by replacing a part of the pipeline. In the Netherlands, the number of leakages on the distribution grid is about 10,000 per year [3]. Given the total length of the infrastructure (some 120,000 km [4]) this is 1 leak per 12 km per year, or framed reversely, a kilometre of grid leaks about once every 12 years. About half the leakages is detected externally (people smelling gas, or third party interventions on the grid), the other half is detected by the gas detection program. This means that spontaneous failures (detected externally) occur once every 25 years/km, and one in five inspections (spaced 5 years apart) per kilometre finds a leak.

However, the grid is not a uniform entity, it consists of many different materials of different ages. Leakage prone materials are cast iron (the joints), asbestos cement (cracking), and steel pipelines (corrosion). As a contrast, the current standard of polyethylene shows much less leaks. Given that the large scale rollout of the gas grid occurred in the 50s and 60s, large parts of the grid are of significant age. Therefore, the time may have come to replace sections of the grid when leaks are found, instead of just repairing the leak. The key question in this consideration is how to determine when that time has come. In this paper, several approaches for making the repair versus replace decision are explored for a typical case.

2 Case Introduction

In a street, a small gas leak has been detected in the gas leak detection program that runs every 5 year. A similar gas leak was detected 5 years early in the same street. There are no records of leaks before that. The pipeline has been constructed in

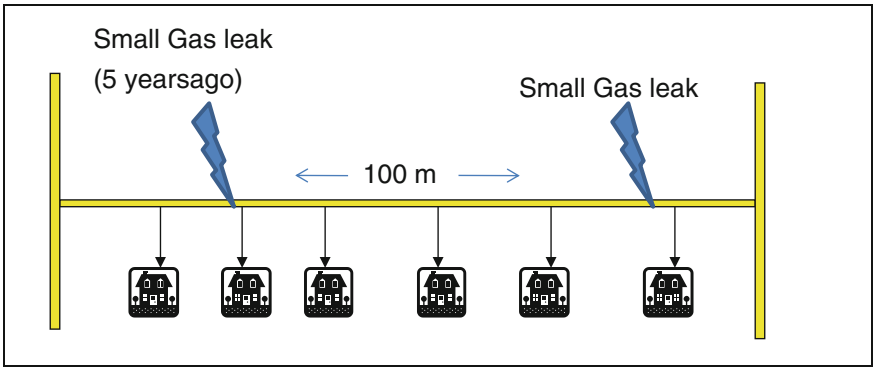


Fig. 1 Schematic representation of case

1953, and may be approaching the end of life. Is it better to repair the leak or to replace the whole pipeline? (Fig. 1).

3 Rule Based Decision Making

In a rule based approach the leak rate would be compared to the replacement criterion. This section of 100 m has a leak every 5 years, which translates into 2 leaks per kilometer per year. Suppose the replacement criterion was 1 leak per kilometer per year (say the top 5 % worst pipelines), the section would pass the mark and would be replaced. A very clear and simple decision. However, it is not evident the criterion properly balances costs and benefits of the replacement. There will always be a worst performing 5 %, but the performance of these worst performers can still be very high if they are compared to an absolute reference.

4 Basic Risk Based Approach

In a risk based approach the decision requires comparing the exposure of doing nothing to the costs of replacing the pipeline right now. Replacing the pipeline costs 5,900 € (50 €/m plus 150 € for reconnecting the houses), whereas the single repair costs 1,000 €. The typical gas leak found in the inspection program is very small, therefore in terms of product loss it can be neglected. However, any gas leak is a safety risk. According to Wijnia and Hermkens [5], the exposure for a small leakage in a main (precursor 13) is on average 177 €. This brings the total cost of a leakage to 1,177 €. The Table 1 summarizes these figures for comparing the options.

However, replacing the asset lasts for 60 years without leaks, and repairing the asset only lasts 5 years. In a 60 year period, this cost would have to be made 12

Table 1 Costs of different options

Leaks	Cost	Replacements	
Cost of leak repair	1,000 €	Replacing pipeline per meter	50 €
Risk of small gas leak	177 €	Reconnecting service line	150 €
		Average number of connections	60/km
Total costs of leak	1,177 €	Total cost per kilometre	59,000 €

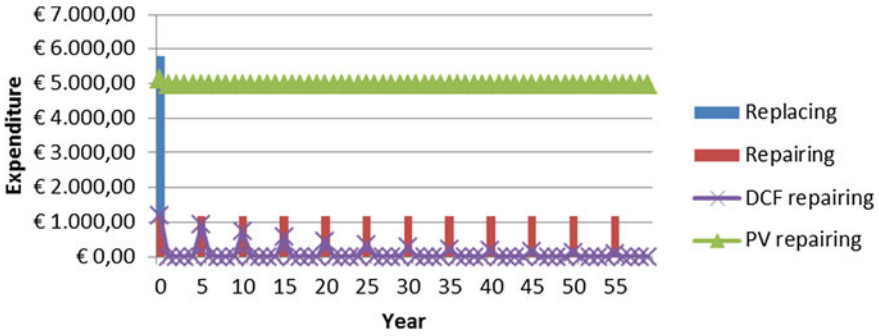


Fig. 2 Comparing replacing and repairing the asset for the next 60 years

times. If for this series of expenditures the net present value is calculated (real interest rate on 5 %), a value of 5,146 € results. See the Fig. 2.

According to this approach, the repair option is cheaper, but the difference is not as big as the initial expenditures may suggest. A more direct approach for comparing different lifetimes is the equivalent annual cost, being:

$$Eq.Ann.Cost = Expenditure * \frac{r}{1 - \frac{1}{(1+r)^N}} \tag{1.1}$$

In this formula, r is the interest rate and N the expected lifetime. Using this formula, the annual equivalent cost are 312 € per year for replacement (5,900 € for 60 years at 5 %), and 272 € per year for repairing (1,177 € for 5 years at 5 %).[1]

5 Accounting for an Increasing Failure Rate

In a more thorough analysis, the difference may even be smaller. The failure rate of assets tends to increase over time, because of wear and tear. The precise development of the failure rate is not known, but some reasonable boundaries can be given. In a study on the long term optimization of asset replacement [6] the failure

[1] The ratio between these two figures is exactly the same as in the NPV calculation.

Table 2 NPV for several failure rate scenarios

	Replacing (€)	Repairing at 0 % growth (€)	Repairing at 5 % growth (€)	Repairing at 10 % growth (€)	Repairing at 15 % growth (€)	Repairing at 20 % growth (€)
NPV	5.900	5.146	12.388	52.422	320.574	2,235.946

rates for a large number of assets were estimated. The development of the failure rate $h(t)$ over time was described with an exponential curve:

$$h(t) = h_0 * e^{\frac{-\ln(h_0)}{T_1} * t} \tag{1.2}$$

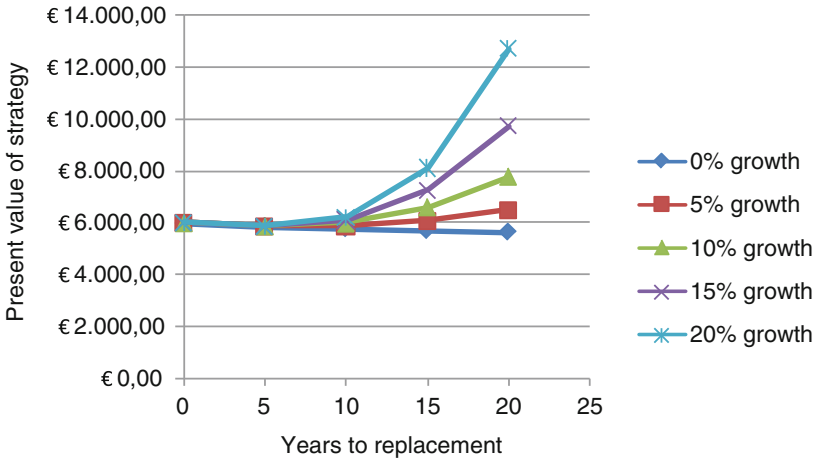
In this formula, h_0 is the failure rate just after commissioning (assuming no infant fatality), and T_1 is the age at which the failure rate becomes 1 per year, that is, the maximum age of the asset. This latter figure was an expert opinion. However, h_0 was difficult to assess directly, therefore an estimate of the trouble free life was established (e.g. the age at which less than 1 % of the population has failed). These two points fix the curve, and h_0 could be calculated. Typical values for $h0$ were in the range of 10^{-3}–10^{-6}, and values for maximum age in the range of 60–100 years. These result in annual growth rates ($=e^{\ln(h0)/t1}$) in the range of 5–25 % per year. Plugging these numbers into the NPV calculation results in the following values (Table 2).

As can be seen, only if the failure rate does not grow, the repair strategy is better than replacing, for all other scenarios it is better to replace the asset.

However, this calculation is stretching reality. If the failure rate increases, it means the inspection results in more leaks to be repaired. But repairing 5 leaks is as expensive as replacing the asset, thus at least if that level is reached, the asset will be replaced. Using the equivalent annual cost, replacement is the best option once 2 leaks are found in the inspection.

Therefore, the proper question is not if to replace the asset, but when to replace it. In the Fig. 3, the net present value for replacement at different (5 year interval) moments is shown.

Replacing the asset right now is for all scenario's regarding the development of the failure rate equally expensive, as is (perhaps surprisingly) replacing the asset in five years. This is because the cost for replacing the asset in 5 years do not depend on the number of leaks found, as the leaks are not repaired. The number of leaks in the next 5 years influences the cost of the alternative to replace the asset later, as can be seen in the rising value of the cost for the scenarios in which the failure rate grows. Based on these numbers, other interesting observations can be made. First of all, the optimal replacement moment for all scenarios except the constant failure rate, the optimal replacement moment is in 5 years. The second observation is that differences are very small. Within any scenario, the costs for replacing now, at 5 years or at 10 years are comparable, i.e. within a 5 % margin of each other. Only at 15 years significant differences appear. The conclusion of this analysis thus is that the asset should be replaced at some moment between now and 10 years, but that it does not really matter when precisely.



	0% growth	5% growth	10% growth	15% growth	20% growth
0	€ 5.900,00	€ 5.900,00	€ 5.900,00	€ 5.900,00	€ 5.900,00
5	€ 5.799,80	€ 5.799,80	€ 5.799,80	€ 5.799,80	€ 5.799,80
10	€ 5.721,30	€ 5.818,25	€ 5.925,13	€ 6.042,67	€ 6.171,63
15	€ 5.659,79	€ 6.053,32	€ 6.561,96	€ 7.218,40	€ 8.063,54
20	€ 5.611,59	€ 6.458,12	€ 7.740,63	€ 9.692,62	€ 12.666,49

Fig. 3 Cost of replacement strategy for different moments at a range of failure rate growth rates

6 Real Option of Postponing the Decision

The analysis above is based on an average: if there is a number of situations to which the above analysis applies, on average replacing the asset between now and 10 years is the best. However, averages do not mean much for individual cases. Furthermore, it does not include the real option value of postponing a decision to get more information. Suppose the decision to replace is delayed until after the next inspection round. If then two leaks are reported, the failure rate most likely grows and it is best to replace the asset, but if there is one or no leaks, the asset can be operated for another 5 year. This can be shown in a decision tree (Fig. 4).

At the current decision moment, the decision can be made to replace the asset, otherwise it has to be repaired. But in 5 years a new decision moment arrives. As has been established earlier, if more than 2 leaks appear, it is better to replace the asset, but if it is only 1 (or zero), it is best to continue operation.

To get an estimate for the option value, the tree above has to be expanded, as there is a cost difference between 1 leak and 2 leaks. Furthermore, as the follow on decision completely depends on the number of leaks, the decision tree is converted into an event tree (Fig. 5).

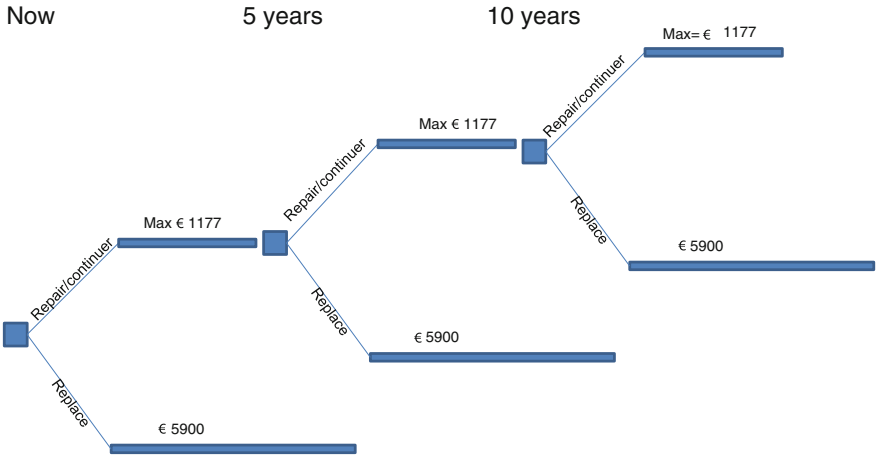


Fig. 4 Decision tree for replacing the asset

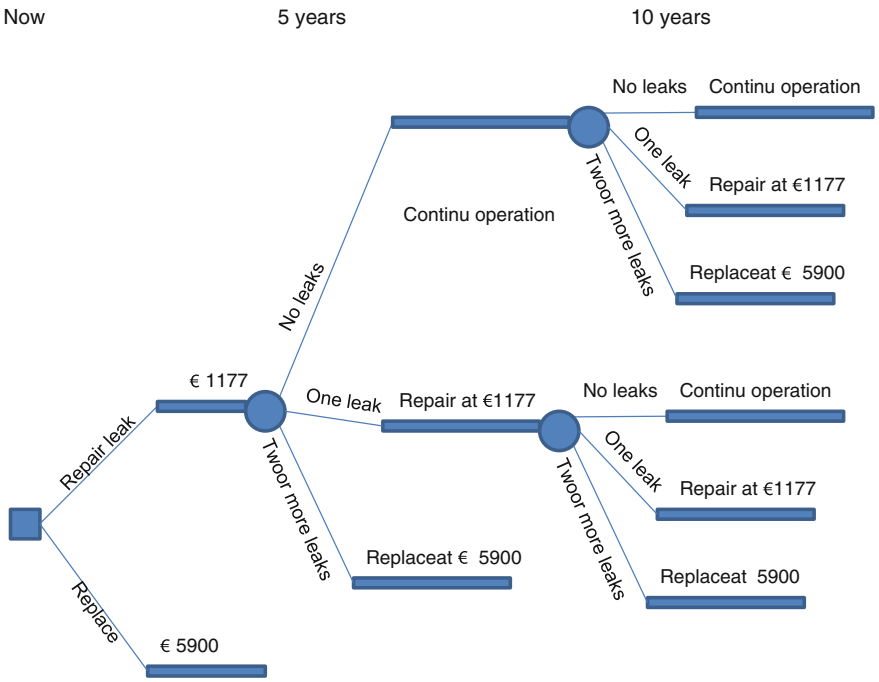


Fig. 5 Expanded decision tree for replacing the asset

If the decision is made to repair the asset, in 5 years' time a new decision has to be made. If there are two or more leaks, the asset will be replaced, if there is one leak, the asset will be repaired at 1,177 €, and if no leaks are found operation will continue for free. The probability that the 5 year period will be survived with a given number of leaks is given by the Poisson distribution. For the scenario with a constant failure rate of 20 % per year, the probability to survive 5 years without leaks is 37 %, the probability of having 1 leak is also 37 %, and the probability of 2 or more leaks is the remaining 36 %. Thus, if the asset is repaired, there is a 36 % probability it will be replaced in 5 years at 6,000 €, a 37 % probability it will be repaired in 5 years, plus a 74 × 36 % probability it will be replaced in 10 years, a 74 × 37 % probability it will be repaired in 10 years, plus a 74 × 74 × 36 % probability it will be replaced in 15 years, and so on. The total value of this decision is given in the following formula:

$$PV = 1177 + \sum_{n=0}^{\infty} \left((0.74)^n \times \left((0.36) \times \frac{5900}{(1+r)^{(n+1)\times 5}} + (0.37) \times \frac{1177}{(1+r)^{(n+1)\times 5}} \right) \right) \quad (1.3)$$

In this formula, n + 1 appears for the discount factor, as the first opportunity to replace the asset is after 5 years. For situations where the failure rate is not a constant, the formula is much more complicated, as the fractions are not constant. It is better to define it recursively.

$$PV(0) = 1177 + (R_0(0) + R_1(0)) \times PV(1) \quad (1.4)$$

$$PV(n) = (1 - R_0(n) - R_1(n)) \times \frac{5900}{(1+r)^{(n)\times 5}} + R_1(n) \times \frac{1177}{(1+r)^{(n)\times 5}} + (R_0(n) + R_1(n))PV(n+1) \quad (1.5)$$

$$PV(N) = \frac{5,900}{(1+r)^{(N)\times 5}} \quad (1.6)$$

This results in the following present values where asset is not replaced as long as it has less than 2 leaks. N is set at 20 (100 years) (Table 3).

As can be seen, there is value in postponing the decision. It is highest for the scenario with a 5 % growth, and lowest for the situation in which the failure rate does not grow. According to the option theory, the option value increases if uncertainty increases, and in a way the 5 % growth scenario has the most uncertainty as it has the broadest optimum (please note this is uncertainty within a scenario, of which it is uncertain it will happen). That postponement of the decision even has value for a 0 % grow scenario may be a surprise, as in that scenario there is no intention of replacing the asset. But if in that scenario two leaks will show up in an inspection, it is better to replace the asset, even though on average it is better to repair.

Table 3 NPV for the strategy to postpone the replacement as long as less than 2 leaks occur, compared to the best option in the classic approach

Scenario strategy	0 % growth (€)	5 % growth (€)	10 % growth (€)	15 % growth (€)	20 % growth (€)
Postponement of decision	4.862	5.231	5.384	5.473	5.534
Replace in 10 years	5.721	5.818	5.925	6.043	6.172
Replace in 5 years	5.799	**5.799**	**5.799**	**5.799**	**5.799**
Replace now	5.900	5.900	5.900	5.900	5.900
Not replacing	**5.146**	12.388	52.422	320.574	2.235.946
Option value	**333**	**548**	**400**	**314**	**256**

7 Conclusion

Summarizing the results above, in a first order approach it was better to repair the asset than to replace it. Adding the risk of an increasing failure rate to the problem, resulted in the best alternative of replacing the asset in 5 years. Including postponement of the decision as an alternative revealed the option value of postponement, resulting in the strategy of repairing the asset as long as less than 2 leaks showed up in inspection and replace otherwise. However, the value differences between all these approaches are limited, especially if the alternative of repairing is omitted. The difference between replacing now and replacing in 5 years is a mere 100 € on 5,900 €, and postponement of the decision adds 5–10 % of value. In other words, replacing the asset somewhere in the next 10 years is a good strategy, as is postponing replacement for as long as less than 2 leaks show up. It may be slightly better, but the difference is hardly large enough to justify the effort of calculation or the effort to convince people they should do something different than they planned. If the standard response would be to repair, that is fine, if the response is to replace the asset that is fine as well. And if there is no standard response, it is better to flip a coin.

References

1. Stewart D, Madsen E (2007) The Texan and Dutch gas: kicking off Europe's energy revolution. Trafford Publishing, Oxford
2. Wikipedia (2013) New London school explosion. (cited 23 June 2013). http://en.wikipedia.org/wiki/New_London_School_explosion
3. Gastec K (2012) Storingsrapportage gasnetten 2011: Apeldoorn
4. ECN (2012) Energietrends 2012. ECN, Petten
5. Wijnia YC, Hermkens RJM (2006) Measuring safety in gas distribution systems. In: First world congress on engineering asset management (1st WCEAM). Brisbane, Australia
6. Wijnia YC, Korn MS, de Jager SY, Herder PM (2006) Long term optimization of asset replacement in energy infrastructures. In: 2006 IEEE conference on systems, man, and cybernetics. Taipei, Taiwan

Assessment of Engineering Asset Management in the Public Sector

Joe Amadi-Echendu

Abstract The importance of asset management in the public sector takes on new significance as capitalism confronts the modern era realities of globalisation and sustainability. The sustainability of environmental, financial and geopolitical systems based on 'real' value demands innovative ways of managing both built environment assets and natural resources. In most countries, the public sector is the custodian of the largest base, especially of infrastructure assets, and with increasing emphasis on accrual accounting, this paper briefly describes a framework for the assessment of engineering assets management in the public sector.

Keywords Public sector asset management · Public sector services · Asset management assessment

1 Introduction

Most people regard an asset as something of worth or value; something useful; a means to an end. These are some of the reasons as to why many would prefer to own and/or manage an asset. In addition to owning, management covers controlling, directing, and supervising the use of an asset toward achieving a profile of desired values. An asset must be fit for the intended purpose, therefore asset management should emanate from organisational strategies, and be consistent with planning, tactical and operational processes. With regard to the public sector, Ref. [1] includes some useful definitions but, as depicted in Fig. 1, the Australian National Audit Office [2] "Better Practice Guide…" provides a useful high level scope of matters that should be considered in the "…strategic and operational management of assets by public sector entities."

J. Amadi-Echendu (✉)
Department of Engineering and Technology Management, University of Pretoria, Pretoria, South Africa
e-mail: joe.amadi-echendu@up.ac.za

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_97

Fig. 1 An asset management framework (ref: State Government of Victoria, Australia)

In general, assets range from natural endowments (e.g., intellect and talent capacities in human beings, land, mineral and other natural resources and the natural environment) to capacities embedded in engineered systems that comprise our built environment. In accordance to the International Public Sector Accounting Standards (IPSAS 17)—"…property, plant and equipment classification recognises separate asset classes such as land, operational buildings, roads, machinery, electricity transmission networks, motor vehicles, office equipment, furniture and fixtures,…" A building, an item of equipment, a machine, a manufacturing or process plant, or any type of physical infrastructure, is typically *engineered* towards satisfying the profile of values as determined by the stakeholders.

An important consideration is who owns or controls the use of an asset. In most countries, the public sector is typically the custodian/owner of the largest base of, especially, infrastructure assets. Psychologically, custodianship/ownership tends to engender a positive attitude, hence it is vital that a manager thinks and behaves like an owner or custodian of the asset. The distinction between public and private sector assets is essentially in terms of custodianship/ownership. This attribute is thus primordial to the way an individual or an organisation manages an asset. The management of assets in the public sector certainly depends on the business objectives of the legal entity called 'public'; and whether or not the public owns, controls, directs, or supervises the use of the asset.

This paper briefly examines the management of engineered assets owned by, or in the custody of the public sector, taking into consideration social, economic and fiscal, and sustainability contexts. The discussion focuses on institutional arrangements for the public sector management of engineering assets intended for the provision of services to the citizenry.

2 Asset Management in the Public Sector

The non-financial assets in the public sector are utilised to provide a wide range of services to the citizenry. These assets include, for example, infrastructure to provide services in education, health, security and transportation, and the services are generally intended to facilitate economic and social development. Although in most jurisdictions, government organisations act as legal monopolies to provide some services, however, the ubiquitous nature of information and communication technologies has provided impetus for increasing trends in internationalisation of public sector services well beyond country-specific monopolies. Thus, on the one hand, the management of engineering assets in the public sector depends on country-specific governance structures, policy, legislative and regulatory regimes, as well as the level of economic, political and social development. On the other hand, external macro-economic and socio-political environments oftentimes exert conflicting and contradictory influences on the management of assets in the public sector. Take for instance, the contradictions embedded in interpretations of multilateral agreements such as the World Trade Organisation (1995) *General Agreement on Trade in Services* (GATS) and the United Nations Agreements on Human Rights. The contradictions tend to be more magnified in underdeveloped/developing country jurisdictions, where it is not uncommon for both the polity and citizenry to argue that certain essential services (e.g., water, housing and sanitation) is a matter of human right and should be provided as 'public good'. The conflict or contradiction arises when the state also has to conform to yet another multilateral agreement such as GATS which essentially promotes *marketization* (i.e., privatisation) of services, of course, not excluding services provided using government owned assets or assets in the custodianship of the public sector.

Arguing from an investment viewpoint, Ref. [3] points out that the advent of privatisation raises the question as to which assets should be owned or controlled by the public sector, and whether or not services should be provided using only government owned assets. They posit that in a mature capital market within a stable socio-political jurisdiction, the management of engineering assets in the public sector would be expected to provide return on the economic investment as well as social dividend as defined by the wishes of the citizenry, albeit that, such wishes are more often bedevilled by contrasting value profiles. With respect to ownership arrangements, public-private partnerships need to be approached from the viewpoint that assets deployed for public service delivery are managed in an effective manner that also provides net return to enterprise, at least recovering both investment and management costs.

Most public entities are becoming subject to financial management and accountability legislation, thus, managers of public sector assets have to concurrently deal with, and contend with

i. fiscal prudence in budgetary execution,
ii. economic efficiency in terms of resource allocation,
iii. social equity in service delivery, and
iv. ecological footprint as a prime sustainability imperative.

Public sector assets constitute a significant item in the governance balance sheet, therefore, the management of land and engineering assets deployed for the provision of services should maximise net benefits to the citizenry. Continuous consideration and harmonisation of social equity demands have to be compared against the utilisation of public sector assets and the fiscal prudence of government agencies. From the sustainability point of view, under utilisation of public sector engineering assets indirectly increases ecological footprint.

The increasing legislative and regulatory adoption of output and outcomes based accrual accounting by nation states also raises the question of capitalisation of public sector assets. Accounting standards generally define public sector assets as resources owned by, and/or in the custody of a legal entity as a result of past investments in potential capacity to provide services that offer social equity, encourage economic efficiency, and promote sustainable livelihoods. International Accounting Standards Board (IASB) defines an asset as "a resource controlled by the enterprise as a result of past events and from which future economic benefits are expected to flow to the enterprise". International Public Sector Accounting Standards Board (IPSASB) defines assets as "resources controlled by an entity as a result of past events and from which future economic benefits are expected to flow to the entity." International Standards Valuation Committee (ISVC) defines public sector assets as "those assets owned and/or controlled by governmental or quasi-governmental entities to provide goods or services to the general public". The focus here is on the ISVC category of operational assets.

Reference [4] contends that accrual accounting capitalisation and reporting of public assets becomes flawed if based on the condition that "… economic benefits or service potential associated with an asset flows to the *owning or custodian* entity…" First, the public entity is the citizenry which is not necessarily homogenous either in economic or socio-political terms. Second, subsequent to this non-homogeneity, the citizenry values often tend to be contradictory, conflicting, contrasting and nebulous, particularly in the context of numerous and widely opposing political dispensations.

Irrespective of the vagarious citizenry values, the important thing is to maximise the utility of public sector assets. Manning in [5] states that maximising public assets "involves a culture shift" that requires integrating the management processes "… into the local growth and regeneration agendas of the community". Samuels in [5] reiterates the importance of having a good register of assets, while Hall and Baber in [5] report that multiplexed utilisation has resulted in savings of 10–30 % in running costs in public asset portfolios.

Whereas it is widely acknowledged that *what gets measured gets done*, this cliché is more applicable to quantifiable economic metrics and financial indicators than to qualitative social equity and political value propositions which are not only fuzzy but also, constantly changing. Furthermore, funding and financing considerations often times provide the impetus for public sector asset managers to manipulate the government fiscal allocation to gain more subsidies, especially when the justification for the investment in an engineering asset is argued from the viewpoint of 'public good' (i.e., for general public consumption as in the case of an educational facility, a health facility, a road network, a recreational park, or even a

weather station). The point here is that for such assets, free users abound and the qualitative value of the asset is very subjective and difficult to measure.

Fair value and market valuation [6] have taken on new significance with the increasing arrangements for public-private partnership ownerships, custodianships and management of assets in the public sector. Irrespective of the method used (cf: [7]), valuation of public sector assets provides an indication of the financial performance and position of government, and thus influences the credit rating of governments and their ability to attract finance from capital markets Notwithstanding the increasing *marketization* of services through public-private partnership arrangements, the contentious aspect of fair value market valuation of public assets arises from the premise that such assets are used to deliver services in a monopoly or absence of any market competition. With the increasing globalisation of previously entrenched local monopolies through the tradability of services, the perplexity from an investment and valuation viewpoint is how to pay, for instance, for the maintenance of an asset that provides public good towards satisfying social equity aspirations.

3 Assessment of Asset Management in the Public Sector

According to [8], the lack of reliable information on public sector assets and systems for their management thereof "…hinders

i. determination of assets' value,
ii. budgeting for asset management activities, and
iii. *evaluation of* public asset portfolio performance…",

thus, resulting in adhoc and reactive attention to planning, utilisation and maintenance of assets in the public sector. They articulate pre-conditions that are necessary for conducting public sector asset management activities efficiently. In addition to the elements shown in the framework in Fig. 1, and the proposals included in the ISO 55000 standard [9], the following factors can be broadly applied to assess public sector asset management (see also Table 1):

Table 1 Perception of public sector asset management

Qualification	Description
Infancy (I)	Tacit or informal planning, lack of systems of audit
Adhoc (A)	Instances of planning, uncoordinated national system of audit
Coordinated (C)	Coordinated planning based on accrual accounting standards
Institutionalised (N)	Planned, auditable financial, legal and regulatory frameworks
Compliant (P)	Clear evidence of compliance to auditable standards

Table 2 A framework for assessing public sector asset management

Factor	Qualification				
Organisational capabilities for management of assets	I	A	C	N	P
Input costs of acquiring public sector assets	I	A	C	N	P
Recurrent budgeting for public sector assets	I	A	C	N	P
Public asset registry and validity of associated data	I	A	C	N	P
Legislative and regulatory policies	I	A	C	N	P
Utilisation of public sector assets	I	A	C	N	P
Performance measurement of asset management	I	A	C	N	P

i. Organisational capabilities for the management of assets in the public sector.
ii. Knowledge of the input costs of acquiring public sector assets.
iii. Recurrent budgeting for public sector assets.
iv. Public asset registry and validity of the associated data.
v. Legislative and regulatory policies for managing public sector assets.
vi. Level of utilisation of public sector assets.
vii. Performance measurement of asset management in the public sector.

Although these factors and qualifications have not been subjected to any scientific scrutiny, however, a high level framework that can be used to assess the management of assets in the public sector may be derived as shown in Table 2.

References

1. Brady WD Jr (2001) Managing fixed assets in the public sector. Universal Publishers
2. Australian National Audit Office (2010) Better practice guide on the strategic and operational management of assets by public sector entities. ISBN No. 0 642 81076 1
3. Brealey RA, Cooper IA, Habib MA (1997) Investment appraisal in the public sector. Oxford Rev Econ Policy 13(4):12–28
4. Christiaens J, Rommel J, Barton A (2006) Should capital assets be recognised in governmental accounting? Paper submitted for presentation at the 4th international conference on accounting, auditing and management in public sector reforms, Siena, 7–9 Sept 2006
5. Guardian Professional (2012) 11.27 BST. www.guardian.co.uk/public-leaders-network/2012/may/24/big-debate-maximising-public-assets. Accessed 27 Nov 2012
6. International Valuation Standards Committee (2006) Exposure draft of proposed international valuation application—valuation of public sector assets for financial reporting
7. Lowe J (2008) Value for money and the valuation of public sector assets. Office of Public Sector Information, HM Treasury, UK. ISBN 978-1-84532-478-0
8. Grubišić M, Nušinović M, Roje G (2009) Towards efficient public sector asset management. Financ Theory Pract 33(3):329–362
9. International Standards Organisation ISO 55000 Asset Management Standards (ibid)

Is Good Governance Conceptualised in Indonesia's State Asset Management Laws?

Diaswati Mardiasmo and Charles Sampford

Abstract Indonesia exemplified its enthusiasm in reforming state asset management policies and practices through the establishment of the Directorate General of State Assets in 2006. The Directorate General of State Assets have stressed the new direction that it is taking state asset management laws through the introduction of Republic of Indonesia Law Number 38 Year 2008; an amended regulation overruling Republic of Indonesia Law Number 6 Year 2006 on Central/Regional Government State Asset Management. Law number 38/2008 aims to further exemplify good governance principles and puts forward a 'the highest and best use of assets' principle in state asset management. However, there is still ambiguity in the meaning of 'the conceptualisation of good governance within state asset management'—particularly in regards to the definition, context, extent, examples, and guidelines. This paper examines state asset management regulations in three Indonesian regional government case studies: DIY Yogyakarta, Gorontalo Province, and DKI Jakarta. This paper introduces the 'Good Governance Evaluator Tool', informed by Miles and Hubermann (1994) work in tabulation and matrix data analysis tool. To facilitate the process of good governance conceptualisation evaluation, it is empirical that each state asset management law, policies, technical guidelines from each regional government is evaluated against the five good governance principles: accountability, transparency, efficiency, stakeholder participation, and regulatory compliance. Through this process which good governance principles are conceptualised, the level in which it is discussed within each clause of a state management law, and the level in which this conceptualisation is understood by state asset managers; can be mapped. This paper emphasises the variance, and at times contradictory nature, in which good governance principles are conceptualised in Indonesia's state asset management laws. As such this paper informs future asset

D. Mardiasmo (✉)
Law and Justice Research Centre, Faculty of Law, QUT, Brisbane, Australia
e-mail: d.mardiasmo@qut.edu.au

C. Sampford
Institute of Ethics Governance and Law, Nathan, Australia
e-mail: c.sampford@griffith.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_98

management policy makers of the quality in which asset governance is exemplified in current laws and technical guidelines.

Keywords Good governance evaluator tool · State asset management laws · Indonesia · Asset governance

1 Introduction

The practice of state asset management is gaining a momentum in importance across governments worldwide [1–4]. Indonesia exemplified its enthusiasm in reforming state asset management policies and practices through the establishment of the Directorate General of State Assets in 2006. The Directorate General of State Assets have stressed the new direction that it is taking in state asset management laws and policies through the introduction of Republic of Indonesia Law Number 38 Year 2008, which is an amended regulation overruling Republic of Indonesia Law Number 6 Year 2006 on Central/Regional Government State Asset Management [5]. Law number 38/2008 aims to further exemplify good governance principles and puts forward a 'the highest and best use of assets' principle in state asset management [5].

This study will focus on analysing state asset management policies in Indonesia, in particular reformed state asset management policies that were introduced with the establishment of the Directorate General of State Assets. Indonesia is chosen as a country case study for several reasons.

Firstly the re-introduction of good governance principles after the Asian Financial Crises in 1997 is an ongoing process within the country, where improving the understanding and implementation of good governance principles remains a constant theme of all presidency regimes after Soeharto and is a main objective of the current presidency regime. Thus there is a push for conceptualising good governance principles in all areas of government responsibilities, including the management of state assets. The Indonesian government and society however acknowledges their tendency to 'remember' the entrenched ways of Soeharto's regime, which was, to a certain extent, contradictory from good governance principles. It is identified that the change in mindset from the familiar 'old' regime to 'new' good governance principle abiding regime is incomplete. Therefore the intricacy of conflicting sets of minds, entrenched ways of doing things, and the optimist objective of conceptualising and implementing good governance principles within public policies provide an interesting platform for understanding the complexities in implementing state asset management reform.

Secondly a review of state asset management practices (of various countries) and the literature on an integrated governance and asset management approach show that although Indonesia is acknowledged to have interesting complexities within its application of reformed state asset management practices [6–8], there is an absence

of studies on Indonesia's state asset management practices—both prior to and after the introduction of state asset management reform in 2006. This shows that there is much to discover, and will have the potential to add to state asset management literature in terms of learning curves for other countries.

Thirdly Indonesia is unique in the sense that it is made up of 33 provinces with different regional cultures, level of resources (human, capital, physical), and government policy objectives. Keeping this in mind, Indonesia introduced decentralisation and regional autonomy regime in 2001, which transfers the authority of governing many sectors (forestry, international trade, etc.) from central government to regional government. This suggests there are potential complexities in the equal implementation of state asset management policies across 33 provinces. A study that focuses on analysing these potential complexities will not only have theoretical contributions—that is within state asset management literature—but also practical contributions for Indonesian state asset management policy makers.

Although the conceptualisation of good governance principles within reformed state asset management policies is suggested, there is still ambiguity on the meaning of 'good governance conceptualisation within state asset management'—particularly in regards to what is meant by good governance principles, which good governance principles are conceptualised, and how these good governance principles are conceptualised. Therefore there is a need for a study that focuses on understanding the relationship between good governance and state asset management laws and policies.

2 The Good Governance Conceptualisation Evaluation Tool

An objective within this study is to evaluate the level of good governance conceptualisation within state asset management laws, policies, and technical guidelines in Indonesia. In order to facilitate the process of answering both questions, it is crucial to perform an evaluation/analysis that compares each state asset management law, policies, and technical guidelines against the five *good governance* principles that are put forward in this study.

The good governance conceptualisation evaluation tool is a table/matrix, aiming to illustrate two goals:

(a) The level of good governance conceptualisation within sections and/or clauses of available (during the study) state asset management laws, policies, and technical guidelines.
(b) The level that an integrated good governance and state asset management approach (which are embodied within each state asset management laws and policies sections and clauses) is understood and implemented by state asset management related actors (i.e. interviewees of this study).

To do so the good governance evaluator tool is set out as a matrix/table that compares both subjects/goals to all five good governance principles through coded symbols.

There are advantages of performing such an evaluation within a table/matrix. First of all the utilisation of a matrix as a comparison tool echoes the benefits of a matrix analysis as identified by [9], where they have recommended its use in qualitative studies as a pattern-identifier tool. Secondly such a matrix assists cross-case analysis, where Eisenhardt [10] has emphasised the importance of identifying similar factors between case studies (i.e. the existence of specific state asset management legal products where applicable) and performing further in-depth analysis by means of pattern-matching logic to draw conclusions. This method of analysis is akin to that of IPC—inferential pattern coding [9], which has been commended to assist cross-case analysis studies, whereby it provides an overall picture of any. Thus the use of a matrix framework provide an overall picture of patterns and interesting comparisons of good governance principles conceptualisation and understanding within intra-regional government and inter-regional government, as well as provide a platform to answer the research questions of the study.

2.1 Coding Rules of the Good Governance Evaluator Tool

The evaluation of good governance principles conceptualisation are based on several coding rules, as the provision of guidelines in how to use the matrix is of great importance [9]. The rules adopted are based on what are deemed to be best/good practices—both in the level of good governance conceptualisation as well as in the implementation stage. Determining factors of best/good practices are taken from asset governance related literature such as that of [11–17]. Based on the available literature on asset governance best practices, as well as state and or public asset management, the following guidelines for the good governance conceptualisation matrix are established and applied:

1. A confirmation of good governance conceptualisation (i.e. a filled circle●) within state asset management legal product is provided on the basis of explicitness. The level of explicitness is defined as 'explicit mention of the good governance principle followed by clear guidelines or advice on how it can be evident during implementation'. Based on the two tier prerequisites within the level of explicitness definition—two levels of coding systems are established. These are:
   (a) One confirmation (●) is awarded if a good governance principle is explicitly mentioned within a particular section or article of the legal product.
   (b) Two confirmations (●●) are awarded if a good governance principle is explicitly mentioned AND clear guideline of how to ensure its evidence through implementation stage is provided.

2. As the confirmation of conceptualisation is based on the level of explicitness, it is therefore important to acknowledge the potential implicitness of good governance conceptualisation (or non-confirmation). Implicitness is further defined as 'implicit or no mention of the good governance principle, however the section or article's content show similarity to the characteristics to a particular good governance principle". Again based on the breakdown of definition for implicitness, two levels of coding systems (Non-filled circles○) are established. These are:

   (c) A non-filled circle (○) is awarded if a good governance principle is NOT explicitly mentioned in the section/article of a legal document, however said section/article contain explanation or implementation guidelines that is of similar characteristics to a good governance principle.
   (d) A cross (×) is awarded if a section and/or article of a legal document does not mention a good governance principle, nor does the content of said section and/or article can be linked to similar characteristics of a good governance principle.

The purpose of the good governance conceptualisation evaluator tool/matrix is not only to evaluate the level of good governance conceptualisation within state asset management legal products, it is also a tool to evaluate the level of understanding and implementation within an integrated good governance and state asset management approach that is evident among state asset management related actors.

It is designed that the left hand-side column and first row of confirmation (filled circle/non-filled circle/cross) corresponds to the level of good governance conceptualisation within legal products, whereas the right hand-side column and second row of confirmation (filled circle/non-filled circle/cross) correspond to evaluate the level of understanding and implementation within an integrated good governance and state asset management approach that is evident among state asset management related actors.

Similar to the establishment of content in the left hand-side column, the right hand-side column content is also drawn upon general information regarding state asset management and the state asset lifecycle as portrayed in any state asset management legal product. The difference here is in the point of view that this content is perceived from, where for the right-hand side column it is necessary to question whether or not good governance principles are understood and evident in the implementation of each state asset lifecycle.

3. Similar to guideline number 1, the confirmation of good governance conceptualisation within state asset management understanding and implementation is also based on explicitness, where explicitness in this criterion is defined as 'ability to explicitly identify or acknowledge a good governance principle within state asset management practices, as well as provide a thorough account of how said good governance principle is evident in a particular activity of state asset management practice'. Again the above definition can be divided into two criteria, as below:

(f) One confirmatory filled circle (●) is awarded if an interviewee is able to identify or acknowledge (the potential) conceptualisation of a good governance principle within state asset management practice, however said interviewee is unable to provide explicit examples or explain how it is evident in the state asset management practice.
(g) Two confirmatory filled circles (●●) are awarded if an interviewee is able to identify or acknowledge a good governance principle within state asset management practice as well as provide explicit examples and explanatory of how it is evident in a particular state asset management practice.
(h) A non filled circle (○) is awarded if an interviewee is unable to explicitly identify good governance principle evidence within a state asset management activity; however said interviewees' account and explanation of a state asset management activity prove to have components that are potentially similar to characteristics of a particular good governance principle.
(i) A filled and non filled circle (●○) mark are awarded if there is explicit mention of the good governance principle during interview however how its conceptualisation and/or evidence in implementation is implied.
(j) A cross (×) is awarded if an interviewee is unable to explicitly identify good governance principle evidence within a state asset management activity, and their account/explanation of said state asset management activity does not relate to any good governance characteristics.

2.2 Example of Good Governance Evaluator Tool

Table 1 provides an example of the Good Governance Evaluator Tool, being applied to an asset management regulation; Gorontalo Provincial Government Governor Regulation 23/2007. For a full list of tables and application of the tool, please refer to Appendix A.

2.3 Justification for the Good Governance Evaluator Matrix

This evaluation matrix is justified in the qualitative methodology analytical tools sense and is established in consideration of the various works in asset management and public/state asset management literature. The matrix is to a certain extent is a pure comparison of state asset management general information and lifecycle against good governance principle; it is not clouded by potential/suspected impediment factors such as culture, organisational differences, resources disparity, or political intricacy. The matrix/tool allows users to simply categorise conformational filled circle/non-filled circle/cross based on document analysis and

Table 1 The good governance evaluator tool/matrix

Central / Regional government	SAM legal product	Section and/or article	Good governance principles					Understanding and implementation
			Transparency	Accountability	Efficiency	Regulatory compliance	Stakeholder participation	
Gorontalo provincial government	Governor regulation 23/2007	Section 1 Articles 1–4. Section 2 Articles 5–6	●	●	●○	●○	○	Definition of state assets and SAM
			○	○	○	○	○	Identification of SAM life-cycle and activities, SAM related factors
		Section 3 Articles 7–10			○	○	○	Planning and budgeting of state assets
			●●	●●	○	●	○	
		Section 4 Articles 11–15	●	○	●	○	○	Acquirement and/or procurement of state assets
			●	●	○	●●	○	
		Section 5 Articles 16–20	○	○		○	○	Storage and distribution of state assets
			○	○				
		Section 6 Articles 21–24		○	○	○	○	Usage and/or allocation of state assets
			○		○	●	○	
		Section 7 Articles 25–30	○	○	○			Administration of state assets, including: recording, inventory, reporting
			●●	●	○	●○	○	
		Section 8 Articles 31–44	○	○	○	○	○	Utilisation of state assets, including: criteria for utilisation and forms of utilisation.
					●○		●○	
		Section 9 Articles 45–49	○	○		○	○	Securing and maintenance of state assets
						●○		

preliminary analysis of interview transcripts. It also allows the user to utilise the tool/matrix as a stepping stone to further in-depth analysis of comparisons and potential reasons for any emerging patterns. Therefore the matrix can be utilised in other asset management related studies, providing that document analysis and interviews are part of the methodology of the study.

3 Findings and Discussion

The objective of formulating Table 1 was to evaluate the level that good governance principles are conceptualised within state asset management laws and policies, and the level that such conceptualisation is understood and implemented.

In regards good governance conceptualisation within the state asset management law, Table 1 suggests variance in three main matters:

(a) The level of good governance conceptualisation within a state asset management law/policy
(b) Which good governance principle is mostly conceptualised within a state asset management law/policy
(c) The level that good governance conceptualisation within state asset management laws and policies are understood and implemented by public policy implementers

Based on the categorisation of confirmation filled circles/non-filled circle as per Sect. 2.1 of this paper and its allocation in Table 1; the categorisation of high, medium, and low level of conceptualised is organised in Table 2.

3.1 Level of Good Governance Conceptualised

The exercise of comparing Table 2 and the good governance evaluator matrix (Table 1) suggest that good governance is conceptualised at varying levels in different clauses and sections within state asset management law, regardless of the SAM law structure. In comparing Tables 2 and 1 it is found that there is medium to high level of good governance conceptualisation in sections that specifically addresses matters such as: (a) planning and budgeting, (b) acquirement and/or procurement of the state asset, (c) reporting of state asset, and (d) the change of ownership process of a state asset.

In regards to the level of good governance conceptualisation within SAM law based on regional groupings, it proved to be quite difficult to categorise, for there is a variance in the level that good governance is conceptualised within each section, chapter, and article within each state asset management law. That said, it is observed that there is an absence of high level good governance conceptualisation in all regional governments that were involved in this study, where a select few are

Table 2 Good governance (GG) conceptualisation and understanding category

Level of conceptualisation and understanding	Conceptualisation category	Understanding category
Low level	X in all five GG principles *or*	X in all five GG principles *or*
	● in one/two GG principles *or*	● in one/two GG principles *or*
	○ In one/two GG principles.	○In one/two GG principles.
Low-medium level	● in some GG principles *or*	● in some GG principles *or*
	○ In some (more than one) GG principles	○ in some GG principles, *or*
		●○ In one/two of GG principles
Medium level	○ In all five GG principles *or*	●○ In some GG principles *or*
	● in all five GG principles *or*	●● in one/two GG principles *AND* ○ in other GG principles
	●● in one/two GG principles	
Medium-high level	●● in some GG principles *and*	●○ In all five GG principles *or*
	○ in other GG principles	●● in some GG principles
High level	●● in all five GG principles	●● in all five GG principles

of the medium-high and medium level category, and most are in the low or low-medium level of good governance conceptualisation. Hence it is preliminary concluded that there is a variance in the degree that good governance is conceptualised within Indonesian regional government state asset management laws. This is as illustrated in Fig. 1.

It is observed that *Regulatory Compliance* principle is most referred to and explicitly mentioned in state asset management laws across all regions. In sections and clauses that explicitly mention regulatory compliance principle there is an emphasis on the need to perform state-asset related tasks and functions (or lifecycle) in accordance to all applicable regulations, laws, and standards—for example according to relevant governor decrees, international standards, or ensuring the correct chain of command is adhered to in the decision making process.

Both *Transparency* and *Accountability* principles are found to be the second most referred to, and explicitly mentioned and explained in state asset management law sections/articles. The mention and explanation of *Transparency* and *Accountability* are mostly found in sections/articles that detail state asset management matters such as reporting, provision of information, and providing accountable justification (including audit of reports) of decisions made (mostly to the regional people's representative) regarding state assets.

Fig. 1 *Good governance* principles conceptualised in Indonesia state asset management laws

Stakeholder Participation and *Efficiency* governance principles conceptualisation is also sighted, and explicit mention of the principle and how to conceptualise both principles are also evident. However there is a tendency for *Stakeholder Participation* and *Efficiency* governance principles to be mentioned in specific parts of the state asset management law—for example *Stakeholder Participation* is mentioned in sections and articles that specifically discusses the change of ownership of state assets only and *Efficiency* tend to be mentioned in sections and articles that specifically discusses utilisation of state assets only. Hence *Stakeholder Participation* and *Efficiency* are not explicitly evident throughout the state asset management law in the manner that *Regulatory Compliance* is. Although *Transparency* and *Accountability* are also found in specific sections only, both governance principles' conceptualisation are found to be explained in depth, whereas *Stakeholder Participation* and *Efficiency* are mostly found to be implicitly adhered to.

In analysing which good governance principle is highly conceptualised in state asset management laws, the finding that *Regulatory Compliance* is most explicitly mentioned and explained is, to a certain extent, a surprise; for based on review of literature concerning Indonesian public policy reform and related state asset management reform, and the opinions of interviewee's, there seems to be a heavier emphasis on the conceptualisation of *Transparency* and *Accountability* good governance principles. This suggests the potential misunderstanding between the perception of interviewees and the 'black and white' content of state asset management laws. Such a misunderstanding can be explained by the thoughts of Mardiasmo [18]; where is a high level of hard control (laws, regulations, etc.) in governing public policy reform, but low level of soft control (educating public policy implementers,

workshops, etc.) to ensure implementation of the public policy reform. The misunderstanding can be explained in two ways: *firstly* there is a high level of hard control—hence regulatory compliance is automatically emphasised in state asset management laws and policies, and *secondly* there is low level of soft control which leads to interviewees incorrectly identifying transparency and accountability as a main factor in good governance conceptualisation.

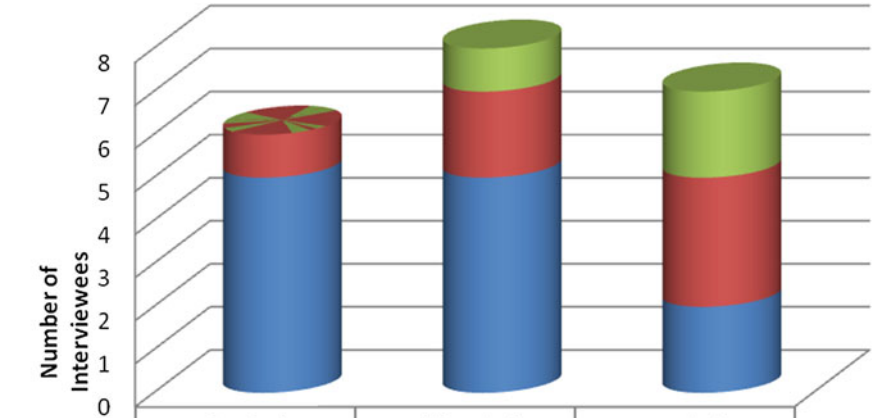
3.2 Understanding of Good Governance Conceptualised Within SAM Laws

The *first interesting observation* is that there seems to be *a discrepancy in the explicit conceptualisation of a good governance principle and the level that it is understood and implemented*, where an explicit conceptualisation does not seem to guarantee a high level of understanding and implementation. An example is DKI Jakarta, where explicit conceptualisation and in-depth explanation of accountability and stakeholder participation are identified within sections and articles of SAM law; yet interviewees do not exemplify high level understanding, with discussions on accountability and stakeholder participation being limited to only acknowledging the existence of the term within the state asset management law.

Interviewees explained this by highlighting the low level of state asset management reform training and knowledge that government official receives; whereby those at high level government (echelon 1) are identified to have more opportunities to gain knowledge than those at middle or lower level of government structure (echelon 2 or echelon 3). As there is an abundance of middle and lower government echelons compared to high level government echelons the discrepancy in state asset management knowledge becomes more obvious/evident within a regional state asset management body/division. This is illustrated in Fig. 2.

A *second interesting observation* is the varying levels of good governance understanding within the state asset lifecycle. Based on the comparison exercise of Tables 1 and 2, the variance in the level that good governance is conceptualised within state asset lifecycle is understood and implemented are as below:

(a) **Higher level of good governance understanding** and implementation in the **early stages of state asset lifecycle** (in particular planning and budgeting; and procurement or acquisition of state assets) and the **end/reporting stage of state asset lifecycle** (i.e. financial reporting and/or inventory reporting of state asset).
(b) Some/**mid level understanding of good governance** principles in the **change of ownership/handover of state assets stage**
(c) **Low understanding** and implementation of good governance principles in the **middle state asset lifecycle stages** such as storage and/or distribution, allocation and utilisation, and securing and maintenance of state assets.



	High Echelon	Middle Echelon	Low Echelon
Low Und	0	1	2
Mid Und	1	2	3
High Und	5	5	2

Fig. 2 Good governance understanding based on echelon level

The observations above provide supporting evidence to the 'lack of caring culture' (in state assets) argument provided by Pardiman [19] and Hadiyanto [5] in their explanation of a stagnant state asset management reform, where there is more of a focus on ensuring that the government has provided a fair, transparent, and accountable procurement process to the public; as well as ensuring that financial reports and inventories are up to date and complete in the event of an audit or a request for report from bodies of higher authority—rather than ensuring maximised utilisation and maintenance scheduling of state assets. Furthermore the observations above also supports the theory that there is an incomplete sense of ownership or stewardship towards state assets, as put forward by Kaganova and Peterson [11], whereby the main challenge in state asset management reform is to change the perception of those who manage it -from perceiving state assets as a free good to one where state assets are a source of wealth, hence more emphasis on maximising utilisation and maintenance.

A ***third interesting observation*** is the varying levels of understanding and implementation for different good governance principles. Based on the number of filled and non-filled circle in the good governance evaluator matrix (Table 2), it can be concluded that:

(a) Higher level of understanding and implementation in transparency, accountability, and regulation compliance principles
(b) Some/mid level of understanding and implementation in efficiency and stakeholder participation principles

It is interesting to see efficiency not highly understood and implemented, for integrated governance and asset management initiatives such as those introduced by Cornish and Morton [12], Kortelainen and Kommonen [14], and Woodhouse [13, 19] emphasised the need to articulate and conceptualise efficiency in state asset management practices. It is even more surprising if compared to the main objective of the Directorate General of State Asset's motto, whereby they have pushed the notion of 'highest and best use of assets' [5], which implicitly emphasises the need for efficiency in asset management. To a certain extent it is not surprising to find transparency and accountability as highly understood and implemented, as the two good governance principles were also the two most highly understood in previous research regarding good governance understanding and implementation within government practices (see Mardiasmo [20]; Mardiasmo Barnes Sakurai [21]); as well as it being explicitly mentioned in the Directorate General of State Asset's roadmap to strategic state asset management [5]. The contrast in stakeholder participation principle conceptualisation (high) and understanding and implementation (mid-level) prove to be an interesting complexity, in particular as analysis of state asset management laws, regulations, and technical guidelines indicate high level of coordination with third parties and stakeholders—such as other regional governments, central government, builders and contractors, developers, government bodies (external audit body, regional people's representatives for example), and the society itself.

4 Conclusion

The purpose of this paper is to evaluate the level in which good governance principles are conceptualised within state asset management laws, and the level in which these conceptualisations are understood by state asset managers. The simple answer to the question is that it varies, in such a proportion that attempting to find a general pattern and a causal effect linkage proved to be a vain effort. There is variance between:

(a) The national view (i.e. central government view of how good governance should be conceptualised in state asset management laws) and regional government view (at large),
(b) Government bodies within the regional government (province, regency, and city government level)
(c) Between regional governments, and also between the different echelon levels within a government body.

Understanding of good governance conceptualisation within state asset management laws and policies is also evident at varying levels, in particular between (a) different echelon levels of a regional government, and (b) between regional government and central government officials. This study shows an opposite level of understanding—the highest percentage at regional government level (40 %) have

low level understanding whereas the highest percentage at central government level (57 %) have high level of understanding. This explains, and lends to, the mismatch in state asset management implementation level between the central government and regional government. The central government, due to their high level of understanding expects high level of state asset management implementation—whereby they have projected their understanding and abilities on the subject of state asset management to regional government officials. On the other hand, regional government officials are challenged in implementing state asset management laws and policies due to their low level of understanding—however this may not be realised by the central government, sparking issues and tension between the two levels of government.

References

1. Kaganova O, McKellar J (2006) sharing experiences—what we do and do not know. In: Kaganova O, McKellar J (eds) Managing government property assets: international experiences. The Urban Institute Press, Washington DC
2. Conway F (2006) Federal asset management in Australia. In: Kaganova O, McKellar J (eds) Managing government property assets: international experiences. The Urban Institute Press, Washington DC
3. Dow P, Gillies I, Nichols G, Polen S (2006) New Zealand: state real property asset management. In: Kaganova O, McKellar J (eds) Managing government property assets: international experiences. The Urban Institute Press, Washington DC
4. McKellar J (2006) Alternative delivery models: the special purpose corporation in Canada. In: Kaganova O, McKellar J (eds) Managing government property assets: international experiences. The Urban InstitutePress, Washington DC
5. Hadiyanto (2009) Strategic asset management: Kontribusi Pengelolaan Aset Negara dalam Mewujudkan APBN yang Efektif dan Optimal. In: Abimanyu A, Megantara A (eds) Era Baru Kebijakan Fiskal: Pemikiran, Konsep, dan Implementasi. PT Kompas Media Nusantara, Jakarta
6. MacAndrews C (1998) Improving the management of Indonesia's National Parks: lessons from two case studies. Bull Indones Econ 34(1):121–137
7. MacAndrews C, Saunders L (1997) Conservation and national park financing in Indonesia, natural resources management project occasional paper no 6. Natural Resources Management Project, Jakarta
8. Kaganova O (2006) A need for guidance in countries with emerging markets. In: Kaganova O, McKellar J (eds) Managing government property assets: international experiences. The Urban Institute Press, Washington DC
9. Miles MB, Huberman AM (1984) Qualitative data analysis: a sourcebook of new methods. SAGE Publications Inc, Newbury Park, CA
10. Eisenhardt K (1989) Building theories from case study research. Acad Manag Rev 14 (4):532–550
11. Kaganova O, Tian V, Undeland C (2001) Learning how to be efficient property owners and accountable governments: the case of Kyrgyztan cities. Public Admin Dev 21(4):1–9
12. Cornish N, Morton K (2001) Asset governance—a radically new way to manage distribution networks in a competitive and deregulated market. Paper read at 16th international conference and exhibition on electricity distribution (CIRED 2001), at Amsterdam

13. Woodhouse J (2004) PAS 55: specification for the optimized management of physical infrastructure assets. http://www.iam-uk.org/downloads/PASworkshop.pdf. Accessed 28 Nov
14. Komonen K, Kortelainen H, Räikkönen M (2006) An asset management framework to improve longer term returns on investments in the capital intensive industries. WCEAN, Queensland
15. Mardiasmo D, Tywoniak S, Brown K, Burgess K (2008) Asset management and governance: analysing vehicle fleets in asset-intensive organisations. In: The twelfth annual conference of the international research society for public management (irspm xii), Brisbane
16. Kaganova OZ, Nayyar-Stone R (2000) Municipal real property asset management: an overview of world experience, trends and financial implications. J Real Estate Portf Manag 6 (4):307–326
17. Kasso Z, Pergerne-Szabo P (2004) Asset management in secondary cities. In: Kopanyi M, Wetzel D, Daher SE (eds) Intergovernmental finance in Hungary—a decade of experience 1990–2000. Budapest World Bank Institute and Open Society Institute
18. Mardiasmo (2009) Kebijakan Desentralisasi Fiskal di Era Reformasi: 2005–2008. In: Abimanyu A, Megantara A (eds) Era Baru Kebijakan Fiskal: Pemikiran, Konsep, dan Implementasi. PT Kompas Media Nusantara, Jakarta
19. Pardiman. (2009). Prinsip-Prinsip Barang Milik Negara/Daerah (BMN/D). In Workshop Training of Trainer Latihan Keuangan Daerah dan Kursus Keuangan Daerah Jakarta, Indonesia: Direktorat Jenderal Kekayaan Negara
20. Mardiasmo D (2007) Good governance implementation and international alignment: the case of regional governments in Indonesia. School of Management, Faculty of Business, Queensland University of Technology, Brisbane
21. Mardiasmo D, Barnes P, Sakurai Y (2008) Good governance implementation at regional governments in Indonesia: the challenges. In XXII International Research Society for Public Management (IRSPM). Brisbane

A Pandora Box Effect to State Asset Management Reform in DIY Yogyakarta

Diaswati Mardiasmo and Paul Barnes

Abstract Indonesia's public policy reform in state asset management ignited from the publication of unfavourable quarterly external audit results, in which many regional governments achieved low results. It is therefore interesting to observe that after the introduction of said reform in 2006 there is slow improvement of the quarterly external audit results, inducing increased concerns as to the level in which it new state asset management laws and principles are understood and implemented by regional government (Padirman 2009). Mardiasmo (2012) posed the question of 'what are the influencing factors to implementation of reformed state asset management laws'; in which the 'voices of reason'—bureaucratic culture, political history, and traditional culture—are identified as a potential explanation to stagnancy in reform. The purpose of this paper is to further analyse the potential role of 'voices of reason' in the conceptualisation, introduction, socialisation, and implementation of newly reformed state asset management laws and regulations; with the aim to determine whether or not 'voices of reason' does play a role, and if so, how. Mardiasmo's (2012) work suggested as such, however its validity and in what form does such influence take shape, is not yet known. In achieving the above objective, this paper will provide an in-depth discussion and analysis of one of the main case studies in Mardiasmo's (2012) work, DIY Yogyakarta Special Region, outlining their version of state asset management laws and regulations; and elements that influences the conceptualisation and implementation of said laws. Further this paper will draw upon qualitative data (available laws and reports, interview transcripts, and observation notes) collected during the period of June–July 2010. Through a meta-analysis and thematic approach this paper creates an ethnography of DIY Yogyakarta's journey (thus far) in interpreting and moulding expected international and national standards to sit comfortably within its 'voices of reason'.

D. Mardiasmo (✉)
Law and Justice Research Centre, Faculty of Law, QUT, Brisbane, Australia
e-mail: d.mardiasmo@qut.edu.au

P. Barnes
School of Public Health, QUT, Kelvin Grove, Australia
e-mail: p.barnes@qut.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_99

Keywords Asset management policy · Decentralisation · DIY yogyakarta · Culture · Duality in governing

1 Introduction

Indonesia's public policy reform in state asset management ignited from the publication of unfavourable quarterly external audit results, in which many regional governments achieved low results. It is therefore interesting to observe that after the introduction of said reform in 2006 there is slow improvement of the quarterly external audit results, sparking increased concerns as to the level in which it new state asset management laws and principles are understood and implemented by regional government [1]. Further, a review of state asset management literature has revealed a dearth in research focused on Indonesia, both pre and post reform [2, 3]. This brings forward concerns regarding the level in which governance principles are conceptualised and understood (by policy implementers) within reformed policies; and in particular the factors that influences its implementation.

Mardiasmo [3] explored the above phenomena in an empirical study involving four provincial governments (and thus twelve regional/local governments in total); concluding that there is a variance in the level in which reformed state asset management laws and principles are conceptualised and understood in Indonesian regional governments; in such a proportion that attempting to find a general pattern and a causal effect linkage proved to be a vain effort. Mardiasmo [3] found that variance existed in three ways:

(a) The level in which each reformed state asset management laws and principles are conceptualised,
(b) The central government and the regional government view of how reformed state asset management laws and principles should be conceptualised, and
(c) The level in which its conceptualisation is understood by different echelon/responsibility levels of the regional government organisational structure.

Furthermore, Mardiasmo [3] posed the question of 'what are the influencing factors to implementation of reformed state asset management laws'; in which the 'voices of reason'—bureaucratic culture, political history, and traditional culture—are identified as a potential explanation to stagnancy in reform. A preliminary analysis suggested the possibility of 'voices of reason' as a hindrance in the translation of international standards into national, and further regional, public policy conceptualisation; and ultimately the implementation of said policies.

The purpose of this paper is to further analyse the potential role of 'voices of reason' in the conceptualisation, introduction, socialisation, and implementation of newly reformed state asset management laws and regulations; with the aim to determine whether or not 'voices of reason' does play a role, and if so, how. Mardiasmo's [3] work suggested as such, however its validity and in what form

does such influence take shape, is not yet known. In achieving the above objective, this paper will provide an in-depth discussion and analysis of one of the main case studies in Mardiasmo's [3] work, DIY Yogyakarta Special Region, outlining their version of state asset management laws and regulations; and elements that influences the conceptualisation and implementation of said laws. Further this paper will draw upon qualitative data (available laws and reports, interview transcripts, and observation notes) to create ethnography of DIY Yogyakarta's journey (thus far) in interpreting and moulding expected international and national standards to sit comfortably within its 'voices of reason'.

2 Literature Review

2.1 Voices of Reason in Asset Management

The ownership of property, whether in the public or corporate domain, is founded upon cultural principles embodied in countries' constitutions, laws, regulations, and norms [4]. Even when there is a similarity in culture, for example Canada and the United States [5], property rights are based on very different legal principles [4]. The fundamental culture differences (both ideological and political) regarding the ownership of state asset management have a profound effect on many aspects of state asset management such as the levels of privatisation that governments will entertain, the divide in authority between central and regional /local government, perspective of state asset ownership and stewardship, perspective of state assets as a 'free good', and national budget and financial management of the asset. In regards to the levels of privatisation that governments will entertain for example. China is at one extreme—maintaining strong state controls over all property rights, whereas countries such as USA, Canada, Australia, New Zealand, and the United Kingdom are at the other end of the spectrum—recognising the need to dispose of real property assets that no longer serve a role in delivering government programs and services [4].

Real property has to do with 'rights' and the ability to bundle, alienate, transfer, and dispose of and otherwise control rights of occupancy and use [4]. Property, whether public or private, transcends mere physical attributes and is inextricably linked to culture and society—real property has economic, social, spiritual, and political values, and those that deal with real property must understand these many dimensions and the complexities that they represent. It should be noted that the right of regional governments to hold property is defined in the constitutions of many countries and in the legislation governing lower tier governments. Hence it is only logical to assume that the ability of regional governments to manage state assets differs between countries depending on the level of authority given to them by the central government and the ideology adhered to in terms of the function of a government and how a country should be governed.

With the legal base of Law 22/1999 on Regional Government and Law 25/1999 on Fiscal Balance between the Regions and the Central Indonesian government;

Indonesia has decentralised its government authority to local government level, with the aim of a government that is more responsive and regional executives more accountable for their actions [6]. This indicates a higher responsibility for local government (both provincial and regional) in implementing good governance aspects. A House of Representatives of the Regions (DPD) was established in 2004 to strengthen the voice of the different regional governments in political decision making, and especially the voice of poorer regions [7].

Kaganova et al. [8] made a crucial observation in regards to how unique country conditions can add complexity to the implementation of a state asset management practices. *They observed that political trade-offs will usually trump management decisions, however there is still the role of asset managers to make their political masters aware of the potential consequences of their ultimate choices.* Differences in ideologies, political history, and cultural values may affect the perspective of ownership of state assets. It is further commented that for some countries where the asset manager is also part of the government body (or appointed by a government body) there is increased complexities. One could argue that said asset manager is caught between pushing for increased efficiency and what is best for the society or the asset, and following the political direction that is drafted or imposed on him/her. One could also argue that the clash so far between the society and the government, or the corporate sector and government is that neither party understand the priorities and assumptions made by one another. Hence with the asset manager being established /appointed by the government potentially suggests less conflict in drafting state asset management policies and implementing it in practice.

2.2 Literature Gap

Instances where developing nations have moved towards converging or implementing international standards within their context is well documented [9–19], however instances where such an attempt have resulted in non-optimal results are also abundant; whereby large differences in perception and culture has been identified as a main reason [20]. This is true in the case of public policy reform in Philippines [11, 14], Thailand and Malaysia [9, 12], African countries [21, 22], and Spain [23]. Many researchers have also identified the mismatch between western and eastern values [20], where international standards developed in the west might not be fully applicable to developing nations in the east; due to differences in eastern perception on what the standard is and how it should be applied [21, 22, 24, 25].

Indonesia have a history of adapting international standards such as the ISO standards, international accounting standards, and international auditing standards [26] to name a few. However evaluation of these adaptations—for example in studies by the Asian Development Bank—acknowledges that such alignments are always adapted to Indonesian specific conditions and capability, where not all aspects are complied with and those that are complied with have been re-adjusted. Literature also stressed that international standards concocted by international

institutions mismatch Indonesia's aspirations on a fundamental level. This mismatch is contributed to ignorance of Indonesia's uniqueness in their attempt to standardize the world [27, 28], are based on an "ideal world" assumptions [29, 30], and may not match Indonesia's reform agenda [31–33].

Thus there is a need for research that provides clear and valid influencing factors to non-optimal translation of international standards in a developing nation's public policy reform, where both international institutions and regional government policy makers/implementers understand the role that each influencing factor play. A research that provides a strategic map as a tool in translating international standards within the context of the country's "voices of reason" (i.e. culture, history, etc.) is thus crucial, both from a theoretical and practical point of view. Theory wise it contributes a scholarly tool to an ongoing debate in the intricacies of countries adopting international standards, whereas in the practical sense outcomes of this research enables and answers current dilemmas faced by developing nations.

2.3 DIY Yogyakarta

DIY Yogyakarta is both a civil government and a monarchy, led simultaneously by an individual that holds the title of governor (or state premier) and sultan. As a governor (or state premier) performing civil government duties, civil law is upheld and enforced; whereas as a sultan and head of a monarchy (sultanate), Javanese and Islamic laws and traditions are followed. Therefore in regards to the conceptualisation, introduction, socialisation, and implementation of a new public policies; the juxtaposition of 'voices of reasons' in DIY Yogyakarta provides an interesting journey to observe and analyse; answering the question of whether, and how, does 'voices of reason' influence public policy making and practice.

The government structure and level of regional autonomy adopted by the Yogyakarta government is one where Yogyakarta, despite it special region status, adopts the 'regional government structure template' as prescribed by the central government. This results in a regional government that is made up of a provincial government (Yogyakarta provincial government), regencies (Sleman regency, Gunung Kidul Regency, Kulon Progo regency to name a few), and numerous cities and villages; of equal level of governing rights. DIY Yogyakarta's government structure is reflected in Fig. 1.

3 Methodology, Findings, and Discussion

3.1 Methodology

A regional government is a chosen unit of analysis in this study as one of Indonesia's economic and politic transition policies in 1999 was the introduction of

Fig. 1 Government structure of DIY Yogyakarta

decentralisation and regional autonomy regime (officially introduced in 2001), allowing regional government autonomic authority on the development of its region and allocation of resources [34–40]. Under this regime regional governments have the authority to establish its own set of laws, rules, regulations, and technical guidelines; with the provision of national/federal government laws etc. being consulted as a base and 'umbrella' laws. Furthermore public policy related literature are either published by central government bodies [41, 42], international institutions [43–45], or based on data released by either or both [10, 22, 46, 27, 28, 29, 31, 47, 48, 49].

The data collection process in Yogyakarta special regional government is that of multi-method qualitative case study, which include three methods: document analysis, semi-structured interviews, and on-site observation [50, 51]. A thematic analysis [52] is utilised in this study, in order to map emerging themes in the data. The first data collection is document analysis, involving the collection of relevant state asset management documents such as: legislation, policies, technical guidelines, reports, etc.

An important gap exists with respect to the involvement of regional government officials in the data collection process, leading to questions of reliability and reality in daily regional government practices. Some attempt has been made in involving Indonesian regional government officials in data collection process, such as studies involving North Lampung, Flores Island, Bali, Northern Sulawesi, and Papua [53–59]; however such literature concentrated on one regional government and is based on legislative review and observation only, not direct government official involvement.

One of the main concerns in inviting government officials to participate in this study is ensuring that there is participation of low, middle, and high level government officials. Government officials classification is based on the Indonesian

Table 1 DIY Yogyakarta list of interviews

Regional government	High level govt official	Middle level govt official	Low level govt official
Yogyakarta provincial government	2	2	2
Sleman regency government	1	2	2
Yogyakarta city government	1	2	2
Yogyakarta internal audit government body	1	3	2

government's classification of public service, as outlined by the Ministry of Human Resource Planning; whereby *echelon 1 and 2a-b ranking* are considered to be *high level government officials, echelon 2c-d and 3 ranking* considered being *middle level government*, and *echelon 4 ranking and other contract personnel* are considered to be *low level government*. Table 1 provides an illustration of government officials involved in the interview process.

The last data collection method is on-site observation. Approximately two days was spent in each government office—which include participating in state asset management related discussion and meetings, observing day to day activities of state asset management related division/sub-division/actors, and visitation to several state asset sites. In particular state assets that are the source of 'management tension', such as state-owned housing and buildings located near the Sultanate, was visited.

3.2 Findings

The complexity of a dual governing system in DIY Yogyakarta (of civil government and Sultanate monarchy) is manifested in their state asset management laws, rules, and regulations; exemplifying the complexity in any public policy conceptualisation and implementation within the region. The management of any state asset within DIY Yogyakarta is governed by both the civil regulations and the Javanese monarchical laws (or way of doing things), regardless of the official ownership status of the state asset. Two reasons were provided by DIY Yogyakarta interviewees to explain this statement:

First of all, prior to Indonesia's independence from the Dutch colonisation, Yogyakarta Sultanate kingdom was already in existence and was the ruling government of the jurisdiction. Hence indirectly any state assets within DIY Yogyakarta were once belonging to, or under the jurisdiction of, the Sultanate. Therefore all state assets' management should be conceptualised and implemented in consideration of the Javanese Sultanate laws, as a sign of respect and acknowledgement of the 'rightful owner'.

Second of all, although it can be disputed that any state assets acquired through the regional budget funds or any other legal means (i.e. gifted from another region, gifted from central government, etc.) are acquired through the civil government; it is important to remember that the head of the civil government, the Governor, is also the Sultan. Hence indirectly, respect towards the Governor's origins must be shown, in the form of considering traditional Javanese Sultanate laws and way of doing things.

Yogyakarta provincial government have chosen to establish its own specific set of state asset management laws and regulations. Interviewees identified that the traditional Javanese Sultanate way of doing things is a strong influence in every aspect of governing, in particular as there is a high level of respect and loyalty from the Yogyakarta's society for their king. This has, more often than not, led to confusion and uncertainty in the best action (or practice) to take when making decisions. This experience has led interviewees from the Yogyakarta provincial government to believe that such specific laws are necessary, not only because it exercises their authority in congruence with decentralisation and regional autonomy, but also to mitigate any potential challenges or confusion caused by the duality of governing. Interviewees believe that strict separation of rules and regulations in state asset management is even more crucial, in particular as state asset management is considered to be a new concept to be learned and implemented by Yogyakarta government officials.

Interviewees identified state asset ownership as a major issue in Yogyakarta provincial government; in particular as according to history there is belief that all state assets once belonged to the Sultanate. This challenge is further heightened as DIY Yogyakarta is classified as an 'old' regional government; its establishment date dating back to the 1945s, and yet prior to the introduction of state asset management reform in 2006 there is an absence of a formal state asset management practice. Hence crucial state asset information such as ownership, utilisation, and current condition, are largely undocumented or organised in a systematic manner.

The Yogyakarta provincial government has identified three fixed state assets that are of main concern: state-owned housing, building, and land. The main issue is that that there has always been a silent struggle between the provincial government (civil) and the sultanate regarding ownership and management rights. The 'silent struggle' for state asset ownership between the civil government and the Sultanate is illustrated by interviewees in Box 1.

Box 1 'Silent struggle' of State asset ownership between Civil government and Sultanate

> "…Land, building, homes—these have always been our main problem…because you know, everything belonged to the Sultanate and the monarchy before hand, and so some people think that technically its all still part of the sultanate, but of course its not, and so it becomes confusing who owns what…"
>
> …The Sultan himself may be a bit conflicted at times, that is to say, he is the governor and the sultan, so sometimes he has to think what is best for the province, not saying he is biased or incompetent, but I cant deny that this duality can sometimes be confusing and causes problems…

…The main problem is, we don't really know what we own. Every time we decide on taking an action on one of the state assets we have to be very careful, make sure that its ok with both the national law and sultanate. Sometimes it can get exhausting and confusing, and it means that everything is slowed down…

Yogyakarta's specific state asset management is designed to explicitly provide guidance to state asset management related actors and the society, as well as signalling (the potential of) a more sophisticated state asset management understanding and system. The positive implications of establishing specific state asset management laws is expressed by Yogyakarta provincial government officials in Box 2:

Box 2 Positive Opinions regarding Specific Yogyakarta Provincial Government SAM Laws

…To a certain extent its good, because you know, we know exactly which laws to refer to, these are the steps that needs to be taken, this is who the document goes to, etc.
"…I like it, because it gives me an exact structure where I know what is the next stage, and I know which law to go to when I need to consult things, and these laws are much shorter—they are about 2 or 3 pages, not the usual book length, so its very easy to read, understand, and implement…"

Although there are positive implications to the key message of Yogyakarta's specific state asset management laws, there are also negative implications. The first negative implication is *the intricate and perhaps complex, multi-layered checks and balances processes* that may prove to be more of a disadvantage. The check and balances process is illustrated in Fig. 2.

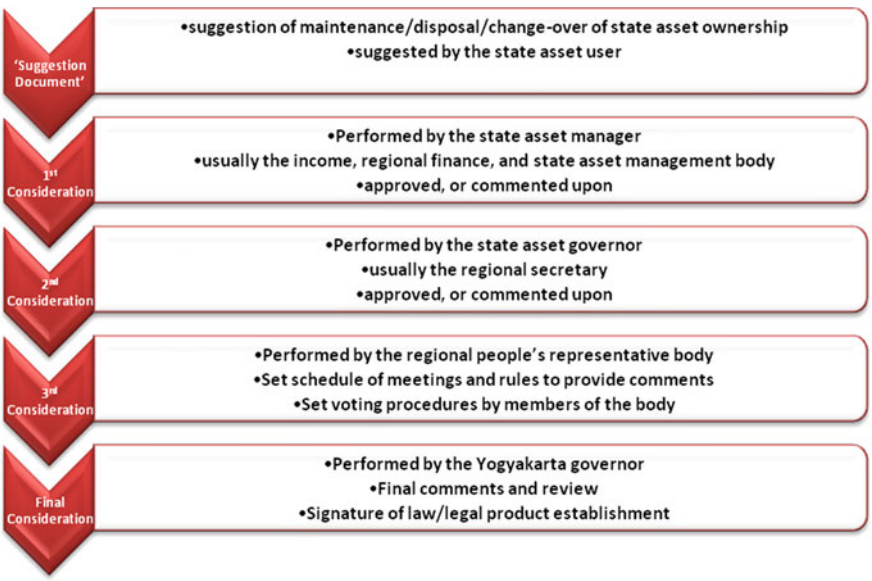


Fig. 2 Processes of legal product establishment in Yogyakarta provincial government

Figure 2 provide a step-by-step process that a state-asset management related document (in particular maintenance, disposal, and change of ownership) must go through, where the length of time taken to complete the process is dependent upon factors such as: availability of government officials, comments and feedback given to the document, meeting schedule of the regional people's representative, and the inclusion of any suggested revisions. Furthermore, the prolonged process of law and/or legal product establishment in Yogyakarta Provincial Government is perceived by some—both officials of the government and the internal audit body—as a sign of *red tape in bureaucracy and inefficient*; as illustrated in Box 3.

Box 4 Negative Opinions regarding Specific Yogyakarta Provincial Government SAM Laws

> …the concept is good, i think, but the thing is, it makes everything so slow and its almost like an excuse for things taking time to happen…for example…well the regional secretary is not in office because he is travelling, or not all members of the regional representative were present for the last meeting so they couldn't take a vote, etc.
> …well it does make sense to have the separate laws because even I'm sometimes confused of whether the house in so and so address is under our jurisdiction or the sultanate, but yes it does create a longer processing time and some times, because it does take a long time, it becomes an 'old case', that is not efficient…

Yogyakarta Provincial Government has identified *uncertainty in state asset related role and task description* as an area of continuous effort and challenge within the process of implementing state asset management reform. Such a challenge is deemed justified by Yogyakarta Provincial Government officials as state asset related roles were non-existent until the introduction of the reform (in the year 2006). Hence the introduction of state asset related roles and positions is considered to be a new addition to the government body and organisational structure, one that needs to be fully embraced and understood, yet is addressed with limited—almost non-existent—experience.

3.3 Analysis

Political history plays a role in the *conflicting loyalty and sense of stewardship* that is at times evident in not only the DIY Yogyakarta government officials and/or state asset related actors, but also in the society. Historically, all citizens or residents of DIY Yogyakarta were part of a kingdom—Sultanate (prior to Dutch colonisation), and are thus loyal to their king and the traditional rituals or ways of the Sultanate. DIY Yogyakarta government officials have identified that many within the governing body subscribes to the perception and ideology of being a royal subject, mainly due to the fact that the their leader is of dual identity: civil government governor and Sultan of the Sultanate. Thus their sense of stewardship is that of 'serving my king', with attributes of high loyalty and the need to consider his 'ways' and 'wishes' in day to day governing. Such views of stewardship is further expedited by Indonesia's high collectivist, high power distance, and high

uncertainty avoidance [5] society, in which there is a higher tendency to act as a group and follow the lead of their leader.

Further, a main challenge that is uniformly felt and expressed by DIY Yogyakarta government officials is the *ownership and management rights of a state asset*. The political history of DIY Yogyakarta highlights that prior to the Dutch colonisation Yogyakarta was a kingdom who owned and managed all within the jurisdiction (land). With the continuation of the Sultanate line after the Indonesian Independence Day, there is still a perception that, to a degree, and by right, the Sultanate owns the lands and everything within it, and not the civil government. This is further strengthened by the introduction of decentralisation and regional autonomy regime in 2001, for the regime, after 32 years of centralised Soeharto's regime, re-introduced the feeling of 'sovereignty' and 'our right' within regional governments and an incomplete reporting system of state assets within the jurisdiction up between 1945 and 2006. That said DIY Yogyakarta government officials can not dispute the fact that acquisition, maintenance, etc. of state assets are partially (if not fully) funded through the DAU—allocated regional government funding transferred from the federal/central government to regional government. As a result, DIY Yogyakarta government officials are more often than not uncertain of who has the ownership rights of a state asset, whether it is full or partial (and if partial, what is the percentage of ownership and accountability actions required), and who has the management responsibilities (again, full versus partial) of a state asset.

DIY Yogyakarta regional government officials have expressed their uncertainty and frustrations in this matter, for there are high ambiguity of 'who owns what and who manages what'. Such frustration is further evidenced in government officials' questions of 'which law should be applied?' for government officials are more often than not conflicted between applying civil government law or the traditional Javanese way of doing things. The decision of which law to implement has the potential for different outcomes and igniting different reactions from the society. Therefore the state asset related actor, due to its high uncertainty avoidance nature [5], is more often than not uncertain about which path to take; which causes lengthy processes in state asset management matters. A main example of this is state-owned houses, in particular regarding the utilisation and disposal stage of the lifecycle.

The Yogyakarta provincial civil government law for state-owned housing clearly states the guidelines of who have the rights to living in a state-owned house and the terms relating to maintenance, rent, sanctions, as well as guidelines on when a state-owned house should be 'returned' to the regional government—for example when a regional government is no longer in a particular position, death, etc. Therefore by civil law, if a regional government official is no longer in his or her position due to retirement, relocation, or redundancy; or has suffered from death, then the state-owned house that they, and perhaps also their family/families, have lived in needs to be 'returned' to the regional government; so that it can be utilised in other ways.

Although the above may make sense in the civil law, and in state asset management practices, it is in fact considered to be 'cruelty' and 'disrespectful', for it breaks, or is in conflict, with one of the oldest, most upheld *Javanese tradition*

known: the responsibility of looking after 'family'. It is believed by the DIY Yogyakarta society that all of Yogyakarta's citizens are part of (or are subjects of) the Sultanate, which means that there is a belief of 'one big family' and the need to ensure prosperity of every member of the family. Hence not allowing the retired regional government official a home, or him/her having to return the home he/she has lived in for all the time she/he has served the government/sultanate/community, is deemed to be inappropriate behaviour.

The same view/belief is upheld when there is a regional government official's death (in particular senior government officials) and their immediate family is, under civil law, pressured to return a state-owned house to the regional government. It is deemed to be 'cruel' to ask the family to give up the house they have lived in, government officials are questioned in terms of their duty to 'look after the society' and their social responsibilities should they pressure the family to return the house.

Another difference is in the acquirement stage of the state asset management lifecycle, where the civil law view takes into consideration, predominantly, economic cost and utilization capacity. The Javanese view does take economic cost into consideration; however it also takes into account religious or mythical beliefs of the Javanese society. For example, it is believed that there is a straight line connecting the Sultanate's palace, Mount Merapi, and the South Sea; which signifies a direct line of prayers, luck, prosperity, and security. Therefore in Javanese/Monarchy way of doing things, building or purchasing a state asset on or near this 'direct line' will be vetoed. The two examples show that there is a divergence in the manner in which state asset management practices are approached and implemented, however in order to establish a definite divergence/convergence conclusion there is further need for a compare and contrast of the two views in all stages of the state asset management lifecycle as well as in the task and functions of all state-asset related division.

4 Conclusion

A high level of confusion in which state asset management legal products to implement has further complications such as hindrance to best practices in state asset management and further mismanagement or at times neglect, under the belief that it is preferable to not implement any laws and policies for fear of implementing incorrect laws and policies. It is therefore crucial to observe and analyse the voice of reason that regional government employees use in determining the appropriate approach, measures, and action regarding state assets management.

One of the potential 'voices of reason' that a government official (or a government body) might turn to, in the face of confusion, is the traditional beliefs or societal culture that are the main roots of the regional government in question. 45 % of interviewees have suggested that during times of confusion they are more likely to refer to traditional culture or what is considered to be 'the right thing to do' according to societal beliefs and the local culture. A potential challenge with this

line of thinking is, that there are differing traditional beliefs and societal culture between regional governments, as well as between governing entities within a regional government (for example depending on what traditional tribal views the governing body is located in). This potentially brings the complexity in state asset management to a heightened level, for the reverence back to traditional beliefs and social culture suggest further inconsistency between the approaches taken by each regional government body.

It is therefore concluded that the 'voices of reasons' does have an influential role in the adaptation, conceptualisation, and implementation of state asset management laws, regulations, and practices in DIY Yogyakarta. How 'voices of reasons' play a role is starting to become clear in this paper, through exploration of practices and views from DIY Yogyakarta government official interviewees. However the depth in which 'voices of reason' plays a role, in terms of every aspect within the state asset life-cycle, still needs further examination; and will prove to be an interesting exercise. It is also necessary to examine the extent in which DIY Yogyakarta regional governments would potentially revert back to traditional culture and practices in the face of confusion; for doing so in decision making may in fact increase the level of inconsistency in state asset management practices and stagnancy of reform.

References

1. Pardiman (2009) *Prinsip-prinsip barang milik negara/daerah (BMN/D).* Paper presented at the workshop training of trainer Latihan Keuangan Daerah dan Kursus Keuangan Daerah Jakarta
2. Mardiasmo D (2010) Asset management and governance. Paper presented at the public lecture series, Yogyakarta
3. Mardiasmo D (2012) State asset management reform in Indonesia: a wicked problem. Doctorate Thesis, Queensland University of Technology, Brisbane
4. Kaganova O, McKellar J (2006) Sharing experiences—what we do and do not know. In: Kaganova O, McKellar J (eds) Managing government property assets: international experiences. The Urban Institute Press, Washington DC
5. Hofstede G (2001) Culture consequences; comparing values, behaviours, institutions, and organisations across nations, 2nd edn. Sage Publications Inc, California
6. King DY (2003) Political reforms, decentralisation, and political consolidation. Can decentralisation help re-build Indonesia? Georgia State University, Andrew Young School of Policy Studies, Atlanta
7. Dwiyanto A (2003) Governance practices and regional autonomy: evidence from governance and decentralisation survey (GDS) 2002. Gadjah Mada University, Center of Population Studies; and Partnership for Governance in Indonesia, Jakarta
8. Kaganova O, McKellar J, Peterson G (2006) Introduction. In: Kaganova O, McKellar J (eds) Managing government property assets: international experiences. The Urban Institute Press, Washington DC
9. Case W (1994) Elites and regimes in comparative perspective: Indonesia, Thailand, and Malaysia. Governance 7(4):431–460
10. Brinkerhoff JM (2005) Digital diasporas and governance in semi-authoritarian states: the case of the Egyptian Copts. Public Adm Dev 25(3):193

11. Bryant RL (2001) Explaining state-environmental NGO relations in the Philippines and Indonesia. Singap J Trop Geogr 22(1):15–37
12. Connors MK (1999) National good governance: a thailand recovery strategy / civic consciousness / cooperation and community in Rural Thailand: an organisational analysis of participatory rural. J Contemp Asia 29(4):547
13. Edgington DW (2000) Deciding the public good: governance and Civil Society in Japan. Pacific Affairs 73(1):121
14. Guess GM (2005) Comparative decentralization lessons from Pakistan, Indonesia, and the Philippines. Public Adm Rev 65(2):217–230
15. Hintjens H (2000) Comprehending and mastering African conflicts: the search for sustainable peace and good governance. J Dev Stud 36(4):185
16. Li JS (2003) Relation-based versus rule-based governance: an explanation of the East Asian Miracle and Asian Crisis. Rev Int Econ 11(4):651–673
17. Sing M (2003) Governing elites, external events and pro-democratic opposition in Hong Kong (1986-2002). Govern Oppos 38(4):456–478
18. Velayutham S (2003) No shame or guilt: the crisis of governance and accountability in Asian economies. Humanomics 19(1/2):12
19. Wu X (2005) Corporate governance and corruption: a cross-country analysis. Governance 18 (2):151–170
20. Blunt P (1995) Cultural relativism, "good governance" and sustainable human development. Public Adm Dev (1986–1998) 15(1):1
21. Bardill JE (2000) Towards a culture of good governance: the Presidential Review Commission and public service reform in South Africa. Public Adm Dev 20(2):103
22. Carroll DBaD (2001) NGOs and constructive engagement: promoting civil society, good governance and the rule of law in Liberia. Int Politics 38(1):1
23. Fernandez-Fernandez J-L (1999) Ethics and the board of directors in Spain: the Olivencia code of good governance. J Bus Ethics 22(3):233
24. Baratta JP (1999) The international federalist movement: toward global governance. Peace Change 24(3):340–372
25. Griffin K (2004) Globalization and global governance: a reply to the debate. Dev Change 35 (5):1081–1091
26. FCGI (2002) Corporate governance scorecard for Indonesia. Forum on Corporate Governance in Indonesia, Jakarta
27. Prasad BC (2003) Institutional economics and economic development: the theory of property rights, economic development, good governance and the environment. Int J Soc Econ 30(5/ 6):741
28. Remmer KL (2004) Does foreign aid promote the expansion of government? Am J Polit Sci 48(1):77–92
29. Maher I (2002) Competition Law in the international domain: networks as a new form of governance. J Law Soc 29(1):111–136
30. Liu GS (2005) Comparative corporate governance: the experience between China and the UK. Corp Govern Int Rev 13(1):1–4
31. Matsushita M (2004) Governance of International Trade Under World Trade Organization agreements-relationships between World Trade Organization agreements and other trade agreements. J World Trade 38(2):185
32. Morrison J (2004) Legislating for good corporate governance: do we expect too much? J Corp Citizenship 15:121
33. Rivera-Batiz FL (2002) Democracy, governance, and economic growth: theory and evidence. Rev Dev Econ 6(2):225–247
34. Federspiel HM (2005) Regionalism in post-Suharto Indonesia. Contemp Southeast Asia 27 (3):529
35. Silver C (2003) Do the donors have it right? Decentralization and changing local governance in Indonesia. Ann Reg Sci 37(3):421

36. Tambunan M (2000) Indonesia's new challenges and opportunities: blueprint for reform after the economic crisis. East Asia Int Q 18(2):50
37. Devas N (1997) Indonesia: what do we mean by decentralization? Public Adm Dev (1986–1998) 17(3):351
38. Agung AAG (2005) Interaksi Pusat-Daerah Dalam Pembuatan Kebijakan di Indonesia. In high level meeting on strategic policy making in Indonesia, Jakarta, p 9
39. Bjork C (2003) Local responses to decentralization policy in Indonesia. Comp Educ Rev 47 (2):184
40. Mishra SC (2001) History in the making; a systemic transition in Indonesia. In Policy support for sustainable social economic recovery project INS/99/002. United Nations Support Facility for Indonesian Recovery, Jakarta
41. BAPPENAS. (2005). Penerapan Tata Kepemerintahan yang Baik (Good Public Governance in Brief), edited by S. P. K. Nasional. Badan Perencanaan Pembangunan Daerah (State Ministry for National Development Planning), Jakarta
42. KNKG (2006) Konsep Penyempurnaan Pedoman Umum Good Corporate Governance. Komite Nasional Kebijakan Governance, Jakarta
43. ADB (2004) Country governance assessment report: Indonesia. In: Bank AD (ed) Country governance assessment report. Asian Development Bank, Manila
44. BaKTI (2006) The new paradigm: knowledge sharing. BaKTI News; Memahami KTI dengan seksama
45. ILGR (2004) Initiatives for local governance reform: report. Kabupaten governance reform and initiatives program (kgrip) / initiatives for local governance reform (ilgr), edited by I. f. L. G. Reform. Department for International Development, Government of United Kingdom, Jakarta
46. Kiely R (1998) Neo liberalism revised? A critical account of World Bank concepts of good governance and market friendly intervention. Cap Class 64:63
47. Ciborra C (2005) Interpreting e-government and development: efficiency, transparency or governance at a distance? Info Technol People 18(3):260
48. Shafer K (2004) The pattern of aid giving: the impact of good governance on development assistance. SAIS Rev 24(1):199
49. Sripati V (2005) The ombudsman, good governance, and the international human rights system. Human Rights Q 27(3):1137
50. Yin RK (1994) Case study research: design and methods. Sage Publications, Thousand Oaks
51. Hall AL, Rist RC (1999) Integrating multiple qualitative research methods (or avoiding the precariousness of a one-legged stool). Psychol Mark 16(4):291–304
52. Boyatzis RE (1998) Transforming qualitative information: Thematic analysis and code development. Sage Publications Inc, Thousand Oaks
53. Elmhirst R (2001) Resource struggles and the politics of place in North Lampung, Indonesia. Singap J Trop Geogr 22(3):284–306
54. Erb M (2005) Shaping a 'New Manggarai': struggles over culture and tradition in an Eastern Indonesian regency. Asia Pacific Viewpoint 46(3):323–334
55. Hassler M (2005) Global markets, local home-working: governance and inter-firm relationships in the Balinese clothing industry. Geografiska Annaler, Series B: Human Geogr 87(1):31–43
56. Henley D (2002) Population, economy and environment in island Southeast Asia: an historical view with special reference to Northern Sulawesi. Singap J Trop Geogr 23(2):167–206
57. Newhouse D (2005) The persistence of income shocks: evidence from rural Indonesia. Rev Dev Econ 9(3):415–433
58. Williams CP (2005) 'Knowing one's place': gender, mobility and shifting subjectivity in Eastern Indonesia. Glob Netw 5(4):401–417
59. WorldBank (2005) Papua public expenditure analysis overview reports: regional finance and service delivery in Indonesia's most remote region. In: Bank W (ed) Public expenditure analysis: overview reports. World Bank, Jakarta, Indonesia

Asset Management Reform Through Policies, Regulations, and Standards: The Need for 'Soft' Interface

Diaswati Mardiasmo and Jayantha Liyanage

Abstract Asset management practices within a country or a region are under continuous reform, particularly with the introduction of new 'hard controls'—rules, law, regulations, and policies, in various sectors. They are expected to provide the necessary foundation for safety, efficiency and other performance needs as well as the frame conditions for managing assets. To a certain extent they also reflect the expectations of international standards. Despite the availability of abundance of documents, organizations tend to spend much time in reality, for instance on analysing the minute detail of policy and regulation, to ensure the compliance as well as validity in terms of expected results. In the modern roles of technical and operational managers, the efforts involving compliance has become a daunting task due to various conditions and complexities in organizational environments. This sheds the light right on the asset reformation process, particularly in terms of feasibility and adoptability Empirical research of 76 regional government officers, acting as state asset managers, in twelve Indonesian provinces and district governments confirmed this standpoint. It was found that despite the availability of comprehensive set of laws, regulations, and technical guidelines; there is a high level of uncertainty, ambiguity, inconsistency, and ultimately non-compliance in asset management practice in these regional governments. It seems that there are other types of absorbed/embedded complexities that tend to remain implicit in the nature of transforming systems, and thus far has been difficult to interpret due to lack of knowledge or understanding of these hidden interfaces that asset managers may have inherit. For instance, a closer analysis of regional government officers reveals other variables in play: ingrained asset management culture, political history of government, religion, and the capability of the asset manager itself; all of which impact the level in which changes in the system are received, interpreted, processed, and implemented. What this suggests is that there may have been too much focus on

D. Mardiasmo (✉)
Law and Justice Research Centre, QUT, Brisbane, Australia
e-mail: d.mardiasmo@qut.edu.au

J. Liyanage
Centre for Industrial Asset Management, University of Stavanger, Stavanger, Norway
e-mail: j.p.liyanage@uis.no

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_100

perfectly 'mechanising a hard control' through policies, regulations, and standards. This implies little or no focus on core attributes of the system in which implementation takes place, and even more on addressing the needs of those who bear major stakes during as well as after implementation process. Thus, this paper argues that acknowledgement and development of 'softer' measures and instruments are just as important, if not even more so, as 'hard controlled' approach to realize the best benefit of ongoing reforms in Asset management practices.

Keywords Asset management policy · Soft control · Asset reform · Compliance · Systems approach · Resilience · Adaptive change

1 Introduction

Indonesia's public policy reform in state asset management ignited from the publication of unfavourable quarterly external audit results, in which many regional governments achieved low performance scores. At the helm of this reform is the Directorate General of State Assets, acting as state asset manager for the Indonesian central/federal government and initiators of new (and innovative) laws, regulations, and roadmaps towards best practice in state asset management. The Directorate General of State Asset Management have stressed the new direction of Indonesia's state asset management through the introduction of Republic of Indonesia Constitution Number 38 Year 2008, which is an amended regulation overruling Republic of Indonesia Constitution Number 6 Year 2006 on the management of central and regional government assets; and the establishment of a 2006–2010 Roadmap to Strategic Asset Management. Furthermore, under the Asset Management Reform (provincial, district, and local government inclusive), in conjunction with the Decentralisation and Regional autonomy regime, regional governments across Indonesia were encouraged to establish state asset management laws, regulations, and technical guidelines as seen fit. Indonesia's decentralisation and regional autonomy regime provides regional governments with the authority to: (a) apply and implement central/federal government law without changes, (b) apply and implement central/federal government law with changes, (c) create its own laws with consideration of central/federal government law.

Indonesia's experiences paramount the phenomena that Asset management practices within a country or a region are under continuous reform, particularly with the introduction of new 'hard controls'—rules, law, regulations, and policies, in various sectors. The underlying expectation mostly is to establish a reliable foundation for safety, efficiency and other performance needs as well as the frame conditions for managing assets. To a certain extent such hard controls may also reflect the expectations of international standards that underline the use of square frames in institutional context. Despite the availability of abundance of documents, organizations tend to spend much time in reality, for instance on analysing the

minute detail of policy and regulation, to ensure the compliance as well as validity in terms of expected results. These on-going efforts appear to abandon the fact that in the modern roles of technical and operational managers, the efforts involving compliance has become a daunting task due to various conditions and complexities in organizational environments. This sheds the light right on the adequacy of the current asset reformation process and practice, particularly in terms of feasibility and adoptability.

In order to explore this status, some empirical research involving 76 regional government officers, acting as state asset managers, in twelve Indonesian provinces and district governments were conducted recently. The study revealed that despite the availability of comprehensive set of laws, regulations, and technical guidelines; there is a high level of uncertainty; ambiguity, inconsistency, and ultimately non-compliance in asset management practice in these regional governments. Subsequently, attention was drawn to the potential influence of other types of absorbed/embedded complexities. One can argue that there remain various other core attributes, mostly implicit, in the nature of transforming systems; of which can be identified as 'soft controls' to the best practice in state asset management. Thus far it has been difficult to interpret the extent in which 'soft control' play a role in the implementation of best practice, due to lack of knowledge or understanding of these hidden interfaces that asset managers may have inherit. Hence the purpose of this paper to explore the facets that makes 'soft control' and identify its influence and importance in asset management reforms.

2 Methodology

To achieve the objectives of this study it is essential to choose an organisation (i.e. a case study) within Indonesia that has the responsibility of managing state assets. A main variable considered in the implementation of reformed state asset management is decentralisation and regional autonomy regime, as it has the potential to add complexity in ensuring equal understanding and implementation of public policy reform [1]. Decentralisation policy in Indonesia indicates regional independence in terms of policy making and economic autonomy [2–4]. Each region has the independence to enact policies based on the level of benefits reaped, where central government acts as an advisory as well as a control mechanism [5–8].

Regional governments established after the introduction of decentralisation and regional autonomy (in 2001) tend to embrace innovative ways of implementing government procedures, due to dissatisfaction to previous practice (i.e. during Soeharto's regime) in governing [9–11]. Mishra [11] concluded that newer regional governments would be more open to implementing good governance principles in their public policy drafting and implementation as this is perceived as an innovative way of governing.

The number of interviews involved within each case study is outlined in Table 1, categorised by echelon level.

Table 1 Number of interviews

Regional government	High level govt official	Middle level govt official	Low level govt official
DIY Yogyakarta	5	9	8
DKI Jakarta	3	4	4
Gorontalo	7	9	11
Directorate general of state assets	2	4	4

As evident in Table 1, there is a slight discrepancy in the number of government officials interviewed; due to time restrictions, availability of government officials, and willingness of government officials to participate. It is also noted that the number of participants from DKI Jakarta and Directorate General of State Assets may seem low (in comparison to DIY Yogyakarta and Gorontalo). Lower numbers of interview participants can be explained by the centralised government structure and state asset management practice of both case studies, in contrast to the decentralised government structure and state asset management practice of DYI Yogyakarta and Gorontalo.

3 Understanding an Asset and Its Regulating Conditions

An asset has a designated function and a role in any organizational, political, economical, and technological environment. By default it has an intrinsic value that can be seen from various perspectives, for instance in relation to economical as well as social. At the same time it possesses a group of stakeholders whose engagements, interests, and subsequent stakes vary depending on the nature as well as the impact level of this value. Both the inherent value and the underlying stakes are two major defining factors when an asset acquires and delivers its function and role in a system. Over the last few years, the attention on asset management has grown considerably as a modern approach to better capitalize and maximize value creation. With this gaining popularity, efforts are also seen on the horizon to introduce standards, policies, and regulations to streamline asset reforms in various sectors, both in public and private. However, this effort has also begun to reveal many bottlenecks and challenges associated with asset management as well as reform implementation processes.

The success or failure of any dedicated effort can in fact be attributed to two fundamental conditions; i.e. (a) *to what extent the underlying content is well-defined or ill-defined in relation to the context*, and (b) *to what extent a particular resolution has been adopted with due attention to the 'soft' regulating factors in a local setting*. This implies the need to avoid potential pit-falls of a 'rigid mechanical

process' (or 'hard control' so to say). Instead, the efforts need to be 'soft tuned' with a greater degree of knowledge and understanding of contextual attributes. At the same time it has to lead the way through a systemic as well as a systematic process for an adaptive change involving a continuous learning process. This is a matter of time, resources, and quality rather than the speed or mere compliance assurance in the bid to meet a set of specific performance targets or expectations.

In principal, during any dedicated effort for an asset reform a number of specific conditions need to be thoroughly understood and evaluated to define an effective path and an efficient course of action. This covers elements both in macro and micro scales representing the actual environment within which an asset exists and performs. There are some core principles that validate the need for a broader understanding of the actual circumstances, which for instance include;

- *An asset is a dynamic element of a complex macro environment*—an asset can have a complex role in socio-economic and socio-political environment depending on the nature and scale of its stakes in macro context.
- *Asset has an organic life*—an asset undergoes various changes and strategies in various types and forms from time to time. At the same time it acquires a unique life profile of its own with specific culture, norms, practices, etc.
- *An asset comprises a complex configuration of different roles and functions*—managing an asset involves many different roles and functions that are configured within a complex structure. The way in which these roles and functions come to play under varying circumstances can mostly be difficult to foresee due to the influence of complex regulating factors.
- *Managing as well as reforming an asset requires a breadth of human skills and supportive tools*—Assets call for a multi-disciplinary (which some may call cross-disciplinary) approach, indicating the need for a complementary set of skills in various capacities. Owing to the breadth of the knowledge and expertise required it also need effective supportive tools to establish efficient professional collaborative interfaces. This has a direct implication on complex decision processes within and without an asset.
- *Adaptive pace of change and Learning experience*—Managing an asset or reformation process is not a one-off activity. It is a continuous process involving various stages of change in different forms and types, through learning and adaptation.
- *Compliance as well as Resilience*—An asset can be subjected to different operational modes from time to time. As much as compliance is important under controllable conditions it is equally important, if not more, to ensure resilience under abnormal conditions. Latent issues embedded in an asset environment have critical roles to play particularly when resilience comes into force.

The underlying key in a systemic process is not a matter of 'hard control'. Rather it is about the development and implementation of action programs that suits the custom conditions by establishing the right level of awareness and knowledge. This in principal is to ensure the right form of 'hard-soft' interfaces to boost performance outcomes.

4 State Asset Management Reform: The Case in Indonesia

Hadiyanto [12] documented reform in state asset management dating back from 2004, with Constitution number 1 year 2004 on State Treasury as a reform locomotive. However Hadiyanto [12] also commented that activities in state asset management policies and practice is still minimal after the introduction of said constitution on state treasury, where time was devoted more to conceptualizing laws, policies, and technical guidelines for said laws and policies as opposed to practice and/or implementation of the constitution. It is with the introduction of Law no 6 year 2006 on state asset management that state asset management reform in Indonesia 'officially' started [12] where the law broadly discusses the following:

(a) State asset management includes activities such as budgeting, acquisition, utilisation, maintenance and monitoring, valuation, disposal, change/hand over, administration/inventory, control mechanisms, and capacity building function.
(b) Introduction of an asset manager role, where the directorate general of state asset management is named asset manager, in a bid to ensure professionalism in state asset management practices.
(c) Integration of managerial and reporting aspects in state asset management in financial reporting as part of accountability

State asset management policies are further reformed through the establishment of Law 38/2008. Based on the most recent law governing state asset management Hadiyanto [12] define strategic state asset management as the integration of functions such as planning, budgeting, maintenance, information system, disposal, and accountability of state asset management that puts forward the principle of "the highest and best use of assets". This definition is within the corridor of asset management definitions, such as the definition offered by Cagle [13] Komonen, Kortelainen, and Raikkonen [14]; Lin, Gao, Koronios, and Chanana [15], Jabiri, Jaafari, Platfoot, and Gunaratram [16], and others.

The concept of ensuring accountability and the highest and best use of asset is also in line with how Loistl and Petrag [17], Cornish and Morton [18], Woodhouse [19, 20], and Marlow and Burn [21] describe the integration of good governance principles with asset management. This indicates a positive start for Indonesia's state asset management reform, however one needs to be slightly cautious based on past performances of introducing new reforms [22].

The Indonesian government has considered the below points to be crucial in improving current state asset management practices [12]:

(a) Increase in society participation—there is a need for the society to increase their level of custodianship and ownership of state-assets.
(b) Increase the level of coordination and participation between different institutions
(c) Establishment of continuous financial support (that is of consistent amount) from the government to other relevant institutions to finance maintenance of state assets.

(d) Provision of data and information system that is accurate, real, and accessible.
(e) Establish a set of norms, standard, guidelines, and manual within the context of state asset management practices.
(f) Stronger rule of law in the rules and regulations that govern state asset management practices.
(g) Further alignment with good governance principles and incorporation of such principles in state asset management rules and regulations.

5 'Soft Interface': Impeding Variables to Asset Management Reform

Throughout the data collection process of this study, interviewees identified numerous influences that play a role in the conceptualisation and implementation of state asset management reform in Indonesia. These 'impeding influences' are presented in Fig. 1, along with the number of interviewees who has identified it as a contributing challenge to state asset management reform.

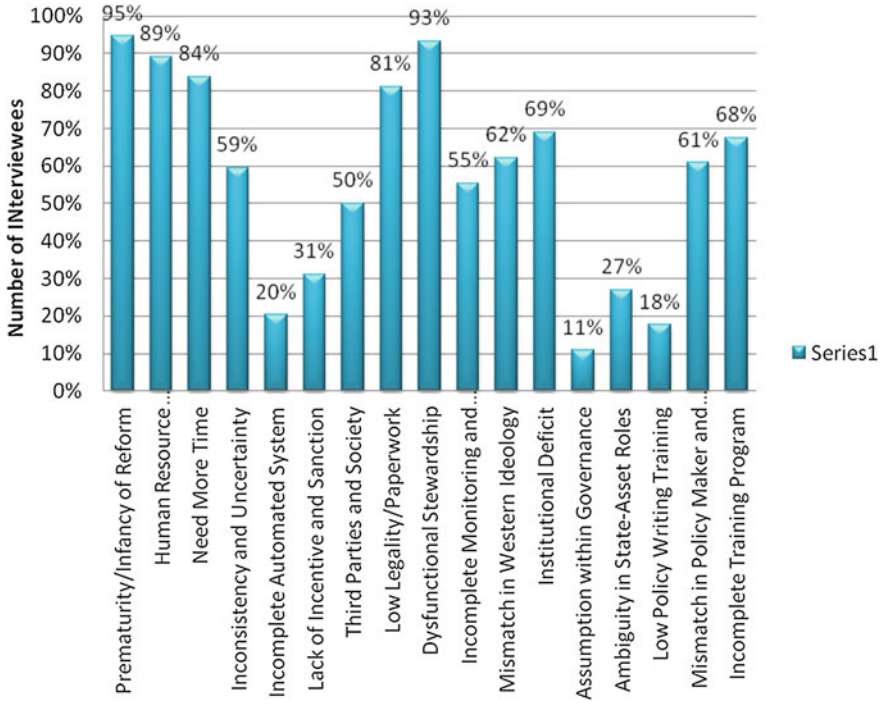


Fig. 1 Support for impeding influences as identified by interviewees

As evident in Fig. 1 interviewees identified five main impeding variables: prematurity/infancy of reform, human resource capacity and capability, the notion of 'I need more time', low legality/paperwork surrounding each asset (i.e. asset data information), and dysfunctional sense of stewardship (towards state assets). It is interesting to observe that out of the five main impeding variables, three are in relation to human factors /'soft control'.

Human resources capacity and capability, the 'People', are identified as an influencing factor to state asset management reform by 89 % of interviewees. Within this influencing factors (human resource) characteristics such as: commitment, level of knowledge, willingness, personality, and capability/competence, and sense of stewardship; are included. In fact, dysfunctional sense of stewardship as an influencing factor is supported by 93 % of interviewees and incomplete training in state asset management laws, policies, and implementation as an influencing factor are supported by 68 % of interviewees.

Interviewees of this study provided several reasoning as to why "the people" is a valid influencing factors. One of them is the disparity in state asset management knowledge between government officials, where it is argued that the disparity in knowledge have resulted in a discrepancy in expectation and outcome. Interestingly, the mismatch between policy maker and implementer gained 61 % of interviewee support (Fig. 1), which suggests that there is agreement surrounding the mismatch as a product of disparity in state asset management knowledge.

Low stewardship and/or dysfunctional perception of custodianship have been identified by interviewees as a potential influencing factor to state asset management. This is supported by 93 % of interviewees, which leads to the belief that it is indeed a valid influencing factors. Here, low level of stewardship or dysfunctional perception of custodianship refers to state asset management actors viewing the state assets in their care as a free good, not a good that needs to be managed in a best practice manner to ensure optimisation of utilisation or wealth creation. State asset management actors do not feel a sense of belonging towards the state assets in their care, viewing that it belongs to another party (though who is unclear) and thus it is not within their interest to ensure best practices in managing the state asset.

6 Conclusion

The increasing importance of an integrated governance and state asset management approach is recognised by various governments around the world. Indonesia is no exception, pledging its adherence to good governance principles within its newly reformed state asset management laws and policies. In 2004 the government of Indonesia introduced a reform in its state asset management laws, focusing on the conceptualisation of good governance principles within policies and practices. This was a response to an incomplete state asset management system, which (prior to the reform) were associated with low state asset utilisation and wealth creation, as well as high costs. However a stagnant and inconsistent implementation of reform is

identified, leading to a reversed outcome of the intended reform decreasing the society's trust towards its government, whom are viewed as guardians to public assets.

A review of Indonesia's asset management reform process suggests an emphasis on 'hard controls', whereby there is a fixation on establishing best practice asset management law, policy, and technical guidelines; ensuring alignment with international standards and/or best practices from other countries. Analysis of impeding variables to asset management reform in Indonesia this approach proved to be a double edged sword. On one side newly appointed regional state asset managers professed uncertainty in asset management practice due to high level of ambiguity in the regulations that govern the practice, calling for clearer and more succinct guidelines that allow them to translate federal and provincial policy into day to day activities/practice. Yet at the same time regional state asset managers have identified three main impeding variables that are not classified as 'hard control', but that of 'soft control': human resource capacity and capability, the notion of 'I need more time', low legality/paperwork surrounding each asset.

This paper highlights the importance of a balanced 'hard control' and 'soft control' emphasis within asset management reform, to ensure that other, ingrained and intangible elements to managing assets are addressed. Particularly so for Indonesia, for their reform in asset management is considered an innovation; with no recorded asset management practice evolution between 1980 and 2004.

The main benefit of this research is in its rich dialogue, exploring the potential of an evolutionary approach to state asset management—one that does not only consider asset management as a theoretical and regulatory form, but also takes into consideration the political history, traditional culture, and human aspects of managing assets. Such a dialogue will assist governments in drafting laws, policies, and practices that are based on consensus; thus reducing the 'disconnect' between public policy makers and implementers. This leads to the reduction of the likelihood of in appropriate state asset utilisation and lower levels of efficiency in state asset management processes/practices; all of which play a crucial role in reducing the high costs currently associated with a stagnant state asset management reform.

References

1. Siddik M (2009) Kebijakan Awal Desentralisasi Fiskal 1999–2004. In: Abimanyu A, Megantara A (eds) Era Baru Kebijakan Fiskal: Pemikiran, Konsep, dan Implementasi. Jakarta, PT Kompas Media Nusantara
2. Devas N (1997) Indonesia: what do we mean by decentralization? Public Adm Dev (1986–1998) 17(3):351
3. Federspiel HM (2005) Regionalism in Post-Suharto Indonesia. Contemp Southeast Asia 27 (3):529
4. Schroeder L (2003) Fiscal decentralization in Southeast Asia. J Public Budg Account Financ Manag 15(3):385

5. Cran R (1995) The practice of regional development: resolving central-local coordination issues in planning and finance. Public Adm Dev (1986–1998) 15(2):139
6. Kristiansen S, Trijono L (2005) Authority and law enforcement: local government reforms and security systems in Indonesia. Contemp Southeast Asia 27(2):236
7. Silver C (2003) Do the donors have it right? Decentralization and changing local governance in Indonesia. Ann Reg Sci 37(3):421
8. Tambunan M (2000) Indonesia's new challenges and opportunities: blueprint for reform after the economic crisis. East Asia Int Q 18(2):50
9. Indrawati SM (2005) Efektivitas Kebijakan Fiskal 2006. Departement Keuangan Republik Indonesia, Jakarta, p 6
10. Kivimaki T, Thorning R (2002) Democratization and regional power sharing in Papua/Irian Jaya. Asian Surv 42(4):651
11. Mishra SC (2005) Pembuatan Kebijakan Demokratis dalam Konteks yang Berubah; Bahan Latar Belakang untuk Laporan Studi Mengenai Proses Pembuatan Kebijakan di Indonesia Discussion Paper Series No. 05/7-IND, United Nations Support Facility for Indonesian Recovery, Jakarta, p 94
12. Hadiyanto (2009) Strategic asset management: Kontribusi Pengelolaan Aset Negara dalam Mewujudkan APBN yang Efektif dan Optimal. In: Abimanyu A, Megantara A (eds) Era Baru Kebijakan Fiskal: Pemikiran, Konsep, dan Implementasi. PT Kompas Media Nusantara, Jakarta
13. Cagle RF (2003) Infrastructure asset management: an emerging direction. AACE Int Trans PM21
14. Komonen K, Kortelainen H, Räikkönen M (2006) An asset management framework to improve longer term returns on investments in the capital intensive industries. In WCEAN, Queensland
15. Lin S, Gao J, Koronios A, Chanana V (2007) Developing a data quality framework for asset management in engineering organisations. Int J Inf Qual 1(1):100–126
16. Jabiri NZ, Jaafari A, Platfoot R, Gunaratram D (2005) Promoting asset management policies by considering OEE in products' TLCC estimation. Engineering management conference proceedings: IEEE international
17. Loistl O, Petrag R (2006) Asset management standards: corporate governance for asset management, 2nd edn. Palgrave Macmillan, Hampshire
18. Cornish N, Morton K (2001b) Asset governance—a radically new way to manage distribution networks in a competitive and deregulated market. 16th international conference and exhibition on electricity distribution (CIRED 2001). Institute of Electrical and Electronics Engineers, Amsterdam
19. Woodhouse J (2004) PAS 55: specification for the optimized management of physical infrastructure assets. http://www.iam-uk.org/downloads/PASworkshop.pdf. Accessed 28 Nov
20. Woodhouse J (2006) Putting the total jigsaw puzzle together: PAS 55 standard for the integrated, optimized management of assets in international maintenance conference. Daytona Beach, Florida
21. Marlow DR, Burn S (2008) Effective use of condition assessment within asset management. Am Water Works Assoc J 100(1):54
22. Lin S, Gao J, Koronios A, Chanana V (2007) Developing a data quality framework for asset management in engineering organisations. Int J Inf Qual 1(1):100–126

Biodiesel Production Status: Are the Present Policies Good Enough for the Growth of Biodiesel Sector in India?

N. Awalgaonkar, S. Tibdewal, V. Singal, J. Mathew and A.K. Karthikeyan

Abstract Worldwide fossil fuel resources such as crude oil, natural gas, coal are decreasing at an alarming rate due to the increasing demand for the fossil fuels across the whole world. More than 75 % of the total crude oil consumed in India is imported from foreign countries. The demand for diesel oil in India is increasing at a staggering rate of 7.5 % per annum. The Indian economy is therefore widely affected by the increasing prices of the diesel oil as most of the heavy vehicle transport in India is run on the diesel engines. So in regards to the present energy crisis and environmental concerns relating to the use and depletion of fossil fuels, the alternative fuel of Biodiesel is gaining importance worldwide as a substitute fuel for diesel in vehicle engines. Due to the increasing dependency of the country on the fossil fuel imports, the Government of India has also made concerned efforts in this regard and launched various ambitious national and state level programmes to promote the biodiesel production and usage in the past decade or so. The recently announced 'National Policy on biofuels' of India (2009) has marked a blending target of 20 % for the biodiesel fuel with that of the conventional diesel fuel by the year 2017. In spite of all these efforts taken by the government in the past decade, the development of the biofuels production sector has been rather slow in India. India is still to attain the 5 % biodiesel blending target that was proposed to be achieved by the year end of the 2010 in the report of National Biofuels Mission in the year 2003. In this chapter literature we have tried to assess the different reasons behind the stunted Biodiesel production sector growth and development in India. The current production potential of biodiesel obtained from major feed-stocks available in India has been studied. The possible impacts of the present national and state level policies on the biodiesel production and usage in different parts of the

N. Awalgaonkar (✉) · V. Singal · J. Mathew · A.K. Karthikeyan
Energy Division, SMBS, VIT University, Vellore, India
e-mail: nimish.awalgaonkar@gmail.com

S. Tibdewal
Chemical Division, SMBS, VIT University, Vellore, India

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_101

country have been assessed. Also, various state as well as national level policy reforms pertaining to the biodiesel production in the country have been suggested, so as to improve the production and usage of biodiesel fuel in the near future.

Keywords Biodiesel · Potential · India · Policies · Drawbacks · Recommendations

1 Introduction

The fossil fuels reserves in India as well as in the world are decreasing at an alarming rate. Consequently, the prices of fossil fuel based energy sources such as gasoline, diesel, furnace oil etc. are increasing exponentially every year. India is heavily dependent on the oil imports for satisfying the energy needs of the country. In the view of rising fossil fuel prices, environmental impacts and climatic changes, different countries all over the world have launched various biofuels programs in order to develop alternatives to the conventional use of fossil fuels. The growth of biofuels production has been enormous in the recent years. The worldwide production of bioethanol rose by 95 % and that of biodiesel increased by 295 % between the years 2000–2005 [1]. Biofuels, as a renewable energy source seems like a feasible option in addressing these energy concerns of the country. The production and use of biofuels in the country like India would help us in issues such as reducing the petroleum imports from other countries, stabilizing the environment, empowering rural people by creating more job opportunities etc.

During these past few years, there has been a growing interest in the field of biodiesel amongst different scientists and researchers. The production of biodiesel from tree borne oilseeds such as Jatropha, Pongamia, Mahua etc. offer an interesting solution in addressing the current energy crisis situation in India. Apart from addressing the energy issues of the country, biodiesel production would help us in creating employment and income for the rural people of the country. It would also help in stimulating the agricultural growth and development in the country. The energy security of the country could be increased. It would also help in reducing the greenhouse gas emissions and air pollution obtained through the burning of conventional diesel fuel. The biodiesel production sector is still in early development stage in India. The government of India has therefore made concerned efforts in promoting the production and use of biodiesel in India. In supporting this new endeavor the government started the work of drafting the National Policy on Biofuels in the year 2003 under the aegis of the Planning commission of India. The National Policy on Biofuels was finally approved in the year 2008 after various debates and modifications in the existing policy draft. The biodiesel production in India is a very different when compared with the biodiesel production scenario in other leading countries. The Indian biodiesel sector is different from biofuel activities in many other countries of the world because biodiesel in India is derived from non-edible oils obtained from oil-bearing trees which can be grown on less

fertile land. Jatropha was identified as the primary important crop for the production of biodiesel in India as it has many advantages over the other oilseeds.

The advantages of using Jatropha are [2]:

1. The oil produced from Jatropha is non-edible, which helps us in eliminating the issue of food-versus-fuel tradeoffs.
2. It has high oil content (50 %) and low gestation period (2–4 years) as compared to the other non-edible oilseeds.
3. It can be grown on the areas of low rainfall and low fertility.
4. It requires less input and minimum care for cultivation.

Energy is been cited as a crucial infix for the socio-economic development as per the National Policy on Biofuels 2008. It also stresses on the abrupt need of substituting petro based fuels by developing biofuels from the inherent renewable feedstock. With the dependency on petro products escalating with each second passing, following objectives were set in two phases by the proposed National Mission on Biodiesel [3]:

1. Phase-I comprised of an Illustrating Project scheduled to be put in action by the year 2006–2007
2. Phase-II was scheduled to ascend in an independent manner for production of adequate biodiesel for the requirement of 20 % blending by the year 2011–2012.

The biodiesel production sector in India has started only a few years ago. Hence, the biodiesel sector in India is still in its primary state, which is mainly due to a lack of economic viability for almost all activities related to the sector. Due to the lack of economic viability for different activities related to the biodiesel production sector in India, this sector is still in its primary stage in India. This chapter critically discusses the following issues:

1. The National Biodiesel Policy along with other union policies and programmes.
2. Various state level policies for the promotion of Biodiesel.
3. Drawbacks in the current Biodiesel promotion policies along with the various reasons associated with it.
4. Various recommendations in policies are suggested to improve the current state of Biodiesel production sector in India.

2 National Biodiesel Policy Along with Other Union Policies and Programmes

India is a federal nation in which the major policy implementation and decision making is vested in the hands of state government under the supervision of national government. Different policy decisions pertaining to agricultural matter, land policies and management, forest management and the rule making for local government are all state matters. Hence different states governments are mostly

responsible for taking major decisions pertaining to the production of biodiesel in India. However, the Union also has important decision making responsibilities such as different tax incentives, fiscal matters etc. Implementation of compulsory or demand side policies, such as compulsory blending of diesel with biodiesel and implementation of various taxes on fossil fuels, tax relaxations on biofuels etc. are all matters of the Union government. The union government has the key role in economic and social development for which it makes the use of various centrally-sponsored Schemes to influence the policymaking in different states. Likewise, one of the decisive field which needs to be chalked out is the Research and Development field, the decisions pertaining to which are mainly addressed by the central government [1].

In the year 2002, a committee on development of biofuels under the chairmanship of the Planning Commission was set up by the government of India. The Biofuel policy was finally approved in year 2008. The national biodiesel policy was not able to gather the required amount of momentum because of the following reasons [2]:

1. Considerable amount of uncertainty was created in the four years between creation and final approval of the Biofuels policy.
2. Farmers and corporate investors had no reliable information as to whether the government would make blending of biodiesel compulsory or recommendatory, what all tax benefits would be available and which crops would be selected.
3. Information available about the Jatropha yields was highly vague. The actual yields of Jatropha plantations were much lower than the actual predicted yields. Jatropha was chosen as the major feedstock for biodiesel production without having necessary research information about the soil conditions, climatic conditions inputs and maintenance activities that are required to obtain economically viable yields from Jatropha. The research findings on the environmental and social impacts of Jatropha plantations were also inadequate to obtain any firm analysis. This can be considered as a major shortcoming of the National Biodiesel Mission's draft. The lack of information on the economics of Jatropha cultivation was also one of the major reasons that affected the launch of policy in a less positive way.

The National Policy on Biofuels firmly states that only biodiesel production from non-edible oilseeds grown on marginal lands would be promoted. The policy also focuses on the fact that oil imports from other oilseeds will not be permitted, rather, oil production from indigenous biodiesel feedstock would be promoted on native land. The policy also clearly states that Biodiesel plantations on government or community lands will be encouraged, while plantation on fertile irrigated lands will be strictly avoided. According to the new policy a number of demand-side support mechanisms would be established as more of the attention would be placed on innovation and research in the concerned biodiesel production areas. A number of supply-side incentives would also be provided for the cultivation of tree based oilseeds, majority of which would not be a part of National Policy on Biofuels [3].

2.1 Demand Related Policies

The National Policy on Biofuels has set the target of blending of petrol and diesel with biofuels to 20 % by the year 2017 [4]. A Minimum Support Price for biodiesel oil seeds is being announced to provide a fair price to the growers. A Minimum Purchase Price for the purchase of biodiesel by the Oil Marketing Companies will also be set by the new policy, and it is fixed by linking it to the prevailing retail diesel price. Although a fixed target for blending fuels has been set, there are still no provisions made to make blending compulsory in the country. Mandatory blending would have been a strong initiation in order to encourage large investments in the tree based oilseeds crop cultivation and transesterification plants for extracting biodiesel from the tree based oilseeds. However it remains to be seen whether the minimum purchase price will be sufficiently high to encourage the production. Already in October 2005, the Ministry of Petroleum and Natural Gas proclaimed a biodiesel purchase policy that came into effect in January 2006. According to the biodiesel purchase policy initiated by Ministry of Petroleum and Natural Gas, oil marketing companies were to purchase biodiesel at a price of now Rs. 26.5 per litre. The oil companies were expected to blend fossil fuel based diesel with biodiesel at a maximum of 5 %. So far, the companies have not been able to produce and procure any biodiesel, as the economically viable quantities of seeds and biodiesel are not yet available and the purchase price offered is much too low for the industry [5].

2.2 Research and Development

The establishment of a sub-committee comprising the department of biotechnology as well as the Ministries of Agriculture, New and Renewable Energy, and Rural Development to support research on biofuels was stipulated by the new biodiesel policy. Already in 2004, the National Oilseeds and Vegetable Oils Development Board (NOVOD) had established a “National Network on Jatropha and Karanja” to contribute in the development of high yielding varieties. Research seems to be concentrated on Jatropha instead of Pongamia or other native tree based oilseeds as the most suitable TBO for biodiesel production. In context with the Current figures, in order to reach economic viability, Jatropha must yield 2 kg of seeds per plant without spending money on irrigation and fertilizers, but however, actual yields under these conditions are well below 1 kg. This focuses on the urgent need for more research and development on the plant material, agro-climatic and soil conditions, inputs, and maintenance activities necessary to obtain high level of productivity of TBOs. Achieving higher productivity is necessary in order to make the industry viable and to increase rural income. Also, there is a considerable lack of research on yielding drought-resistant varieties of different oil-bearing tree species that would give acceptable yields. In this present situation, the assumption that

Table 1 Roles of various ministries involved in the growth of biodiesel sector in India [1]

Ministry	Role
Ministry of new and renewable energy (MNRE)	Research and technological development, overall policy making
Ministry of petroleum and natural gas (MoPNG)	Overall pricing and procurement policy development, marketing
Ministry of agriculture (MoA)	Research and development on the feedstock oilseeds
Ministry of science and technology (MoS&T)	Biotechnological researches on the feedstock oilseeds
Ministry of environment and forests (MoEF)	Implementation of tree based oilseeds on forest wastelands, inspecting the environmental effects of the biofuels

Jatropha and other oil-bearing tree species can be grown profitably on unfertile, barren land does not hold. If drought-resistant tree based oilseeds were available, they would help farmers in having additional income [6] (Table 1).

2.3 Supply-Side Policies

The New Biofuels Policy haven given the 'declared goods status' on biofuels. This means that the biofuels will attract a uniform central sales tax or VAT rate rather than the different sales tax rates prevalent in the states, and movement of the biofuels within and outside any state will not be restricted. The government has already given them the status of a 'non-conventional energy resource', meaning that biofuels would be fully exempted from excise duty. At the current purchase prices, this 'non-conventional energy resource' status reduces the price of biodiesel by about 0.06 $/litre. This price does not, however, outweigh the benefits that conventional diesel enjoys by the heavy subsidies provided. In addition, according to the legal definition the biodiesel is not recognized as a renewable energy source, which would therefore, not allow the investors to obtain additional tax benefits.

In order to encourage the supply of biodiesel, NOVOD also initiated a back-ended credit-linked subsidy programme specifically for TBOs [7]. The program provides subsidies for (a) commercial plantation and nursery raising, (b) establishment of procurement centres and (c) installation of pre-processing and processing equipments. It can be extended to governmental organizations, NGOs etc. Centrally-sponsored schemes are the main elements of the concerned biodiesel policies [8]. There are a large number of centrally-sponsored schemes available than can be used for the growth and promotion of biodiesel plantation. These schemes could be used for the promotion of biodiesel plantation and thereby start the supply of TBOs on a large scale. Some of these centrally sponsored schemes are- National Rural Employment Guarantee Scheme (NREGS), National Afforestation Programme, Village Energy Security Programme, Watershed Development Programme [2].

3 State-Specific Policies

The various state-level policies and incentives supporting Biodiesel production in India are given below [1].

3.1 Rajasthan

As a programme for wasteland reclamation and generation of livelihood opportunities, the state of Rajasthan encourages biofuels planting missions. The responsibility of promoting biofuels in Rajasthan is entrusted to Biofuel Authority of Rajasthan (BFA). The main focus of this mission is on the plantations of Jatropha Curcas on a target area of 21 lakh Hectares of wasteland. Under the Rajasthan Land Revenue (Allotment of Wasteland for Biofuel Plantation and Biofuel based Industrial Processing Unit) Rules 2007, private companies and government enterprises would be provided wastelands on lease up to a period of 20 years. Still, the most pressing problem + that the state continues to face today is the lack of proper processing facilities that can support the emerging processing requirements. Also, the local farming community feels that any significant influence on the livelihood is possible only through introduction of decentralized value-addition options.

3.2 Chhattisgarh

The cultivation of biofuel crops like Jatropha and Pongamia in Chhattisgarh is strongly backed by sufficient land resources and favourable climate. In 2005, the biofuel programme in Chhattisgarh was launched with the creation of Chhattisgarh Biofuel Development Authority (CBDA) under the aegis of Chhattisgarh Renewal Energy Development Authority (CREDA). The forest department and the department of rural development are also strongly active in this mission. The Government of Chhattisgarh has provided provision for private investors for undertaking jatropha/pongamia plantation and setting up biodiesel processing facilities on government lands procured on lease. The government has announced a minimum support price of Rs. 6.50/kg for jatropha seeds and of Rs. 6.00/kg for pongamia seeds so as to ensure fair prices for farmers [9]. A trans-esterification plant of the capacity of 1 kl per day for biodiesel-production from jatropha seeds has been established in Raipur. Rural electrification in a cluster of 50 villages has been achieved by establishing biodiesel based power generators. The government has also plans for setting up a state-of-the art laboratory with a capital outlay of about Rs. 1.5 crores for testing oils, biodiesel, etc.

3.3 Uttarakhand

At the grass root level biofuel promotion in Uttarakhand, is carried out by the Van Panchayats, JFM committees and SHGs. Here, jatropha plantation in Van Panchayat lands, undertaken by the FDC along with the participation of SHGs and JFMs, supplies the seeds to the biofuel processing plant run by the Uttarakhand Biofuel Limited (UBL). A target of covering an area of 2 lakh hectares with Jatropha by the year 2012 was set by UBB. A Jatropha gene bank has been established to preserve the high-yielding varieties. Biofuel-based rural electrification of remote villages in the state is carried out by the board in collaboration with MNRE. The beneficiaries are currently getting a price of Rs. 3.50/kg for Jatropha seeds under the tripartite agreement. In comparison to other states, this price is lower because of lack of competition in the state.

3.4 Orissa

The Biofuel Policy developed by the state of Orissa aims to utilize 30 % of the state's wastelands (0.6 million hectares) and expects to generate 10 million person-days of work through biofuel production in the state. The estimated potential is around a 1,000 kl per annum. According to the policy, Jatropha and Pongamia are the most suitable oil-bearing species which can be chosen for biofuel production. Currently the biofuel programme in Orissa is only at the inception stage.

3.5 Karnataka

Non-edible oil bearing crops like Jatropha, Pongamia, Simaruba, Neem and Mahua has been identified by the government of Karnataka as feed-stocks for biofuel production in the state. Similar to the biofuel policies of other states, this policy also aims at private-public partnership models through long-term lease of government fallow land. In the year 2009 A 'State Task Force on Biofuel' was established to advise the government on policy and programmes related to biofuel from time to time. To oversee the implementation of State Biofuel Policy it is proposed to set up an apex agency by the name Karnataka State Biofuel Development Board (KSBDB).

3.6 Andhra Pradesh

A draft biodiesel policy was introduced by the government of Andhra Pradesh in the year 2005 to facilitate investors and farmers to plant oil bearing trees, mainly Jatropha, Pongamia and Simaruba. The Policy has proposed a tripartite partnership

amongst government; farmers and industry wherein the provision of buy-back arrangements for seeds and credit disbursement for farmers routed through industry are the major highlights. The entire programme was terribly affected when the beneficiary farmers diverted the 90 % subsidy meant for Jatropha irrigation to other crops. Now, the state depends more on Pongamia as a feedstock crop.

3.7 Tamil Nadu

As in the case of Andhra Pradesh, the initial efforts of the government of Tamil Nadu to distribute Jatropha seedlings free of cost to farmers backfired due to lack of maintenance. Presently, the government is taking a cautious approach in promoting biofuel crops. The official Biofuel Policy of Tamil Nadu was released in the year 2007–08 with a target of promoting 100,000 ha of Jatropha plantations over a period of 5 years. On the financial side, Primary Agricultural Cooperative Banks in the state are extending subsidized loans to farmers involved in Jatropha cultivation. The Industrial Policy of Tamil Nadu states that 50 % subsidy is applicable to planting material for Jatropha and other biofuel crops and extends the subsidy available to the agro-processing industry to biofuel extraction plants. To incentivize the processing plants, the government has exempted purchase tax on Jatropha seeds and VAT on Jatropha oil for a period of 10 years from the date of commercial production. The state is actively promoting contract farming for Jatropha cultivation.

4 Drawbacks of the Current Biodiesel Policies and the Recommendations Suggested in Improving Them

The shortcomings in the current national as well as state level Biodiesel policies are discussed in the subsequent section. Various recommendations and suggestions in improving these policies for the growth of Biodiesel sector are also suggested [2].

4.1 Dejected Response from the Farmers

A major hurdle in the way of national bio fuel mission is the acquisition of barren waste land, which can be used for cultivating biofuel feedstock. Majority of wastelands' ownership records do not exist, some are owned by the forestry, some are occupied or owned by marginal farmers and landless labourers and some are under the occupation of rural rich. These waste lands often classified as common property resources (CPR) constitute major portion of rural equity. The day to day necessities of a rural household are met from the resources obtained from these wastelands. Hence, wastelands should be cultivated with plantations that meet the

rural needs viz. fodder for cattle, firewood, fuel wood etc. The committee of national mission is aiming for cultivation of Jatropha Curcas and making it the primary shrub for oil seeds production. However, Jatropha does not seem to fulfill the necessities of local people. The leaves of Jatropha cannot be used as fodder. Also, the yield of wood from Jatropha is very less. Therefore, the local farmers are reluctant to grow Jatropha. Though waste lands are one of India's vast untapped resources, a major part of it is not suitable for cultivation due to overgrazing and overexploitation. Also the change in soil conditions of these lands question the viability of bio fuel plantation. Jatropha Curcas, a perennial shrub which produces oil seeds is capable of withstanding any adverse agro climatic conditions including floods and drought [10]. But these affect their seed production and oil content. Also to obtain higher plantation density the plant requires the soil and water conditions. We are aware of the fact that irrigation plays an important role in increasing yield as well as alternate quality. Most of the perennial plants are capable of withstanding arid and dry condition but they do require proper irrigation during the initial years or else the productivity decreases drastically. Same is the case with Jatropha [11]. Another disadvantage is the lack of small farmers in this field. The uncertainties pertaining in the cultivation of the biofuel crops, the long maturation period and not involving local farmers and communities in decision making related to allocation of CPR are reasons why small farmers do not venture into bio fuel plantations. Also subsidies provided for the cultivation mainly benefit the large farmers and it is found to be higher than the cost of cultivation.

4.2 Sorgum and Caster: Potential Alternatives

Biofuel crops like Jatropha and Pongamia are perennial shrubs and need a long maturation phase. Multipurpose Short duration annual crops are potential alternatives that can yield fodder along with fuel. Through these crops daily needs of small farmers will also be met. This is when annual crops like castor and sorghum come to the rescue. These crops can be used to produce ethanol and biodiesel and allows the farmer to adopt crop rotation. Sorghum can be cultivated in arid and semi-arid regions and it is resistant to adverse agro-climatic conditions like floods, drought and saline-alkaline soil. The main advantage of sweet sorghum compared to Jatropha is the familiar cultivation method. It also produces more bio fuel than Jatropha. It can also be used as fodder for cattle. Higher rate of photosynthesis in these plants are the reason for the higher sugar content in the juice of their stem. This can be processed into sugar, jaggery and ethanol and can be used for production of electricity. So far sweet sorghum is one of the most sought out energy plant in the world. Castor is mainly cultivated for its oil bean. Caster seeds, which are the source of oil, contain 40–60 % oil. The cake obtained after extracting the oil can be used as organic manure. One of major advantages of castor crop over other oil crops is its short maturation period. It can be cultivated in rotation or mixed with other crops.

4.3 Research and Development: Need of the Hour

Research and development department should lay more emphasis on developing Jatropha plants varieties with high yields. It needs expertise of efficient techniques of cultivation and accordingly trains the farmers in different regions. Currently, well established and efficient institutions are very few that are responsible for Jatropha cultivation and therefore, shortage of such institutions was reduced by using NGOs who lack the required knowledge of Jatropha cultivation and thus are unable to train the local farmers as per the needs. In remote areas, NGOs are often outsiders and lack knowledge of local agro-climatic conditions making it difficult for them to choose the desired variety of Jatropha suitable for those areas. Therefore, appropriate institutions are needed that would do tests and experiments along with field trials for the plant varieties developed at the research centres. Demonstration farms also need to be established at various agro-climatic regions that provide training to the farmers. As biodiesel production is aimed for good number of decades, sufficient efforts are required to set up the standards of plant varieties and organization for transfer of knowledge and technology to the farmers. Also it is important to realize the need for various set ups like self-help groups, civil society and cooperatives. Solutions to complications related to use of biofuel in automobile industry like poor lubrication, problem in cold starting due to increased viscosity at low temperatures, increased NO_x emissions because of very high temperature in combustion chamber and oxidation of biofuel during storage needs high quality research [12, 13]. Modifications in injection timings and durations of petro-diesel engines need to be addressed to improve combustion of biofuels [14]. Exhaust gas recirculation results in NO_x emissions reduction by around 50 and 15 % reduction in smoke emissions [15].

4.4 Financial Factors

Biodiesel Association of India demanded an increase in price of Jatropha biodiesel from Rs. 26.25 to Rs. 36 per litre. But in current scenario, several literatures report the tentative pricing of Jatropha biodiesel to be 46.45 per litre. This is the case after omitting the taxes to-be-levied and the money margins for investors/stakeholders. This makes pricing an important factor in the production of biodiesel. But no subsidy has been announced by the Government of India on biodiesel as yet. An elementary push to the Jatropha biodiesel project can be given by lending initial financial aid by the Government until the economics auto-accepts the biodiesel. Further, government should incur for the duty losses faced by the state during campaigning of any biofuel as no support by the government would discourage the involvement of the states. On the other hand, a significant part of a state's revenues is raised from the taxations on petro-based products. Insurance cover schemes which serves as a confidence builder for investors and small local farmers were absent in the first phase of National Biodiesel Mission itself. This is one of the

major reasons for failure of the mission in first phase with no returns on biodiesel production. Some insurance policies exists that cover cost of replanting and loss of income. Hence, schemes that insure yield risk must be introduced by the insurance companies in public and private sector.

4.5 Other Recommendations

Cultivation of Jatropha also ensures large scale rural employment for a large share of unemployed or under employed Indian population. Also, 62 % of the expenditure on Jatropha is in the form of direct wages for unskilled labor. Thus any biofuel promotion policy is an opportunity for mass employment and schemes like MNREGA can also be incorporated with the biofuel promotion policy. Efforts must be taken to promote biodiesel as a "renewable energy fuel". It would also give tax relaxations to the stakeholders, thus, creating and maintaining interests among them. Environmental taxations should be imposed on fossil fuel based automobiles. Such taxations would turn the heads in favour of renewable sources of energy. Although there is no such bill passed by the parliament of India which authorizes Environment related taxations, other ways such as mandating blending of petro-based diesel with biodiesel would prove to be solid. Usage of biodiesel in Public and governmental carriers such as trucks, buses, railways etc. and other large scale petro-fuel consuming organizations should be motivated and appreciated. Pricing of food items must be closely observed as production of Jatropha biodiesel requires wastelands which itself are home for a variety of plants, crops, trees etc. As the wastelands would get continued to be used for biodiesel production, food supply may go down, consequently raising the demand and causing price inflations in food items. Thus the level of blending should be escalated in a planned and organized way rather than sudden changes. Exports on Biodiesel should be allowed, if the product is able to get a big demand and consequent price in the global sectors. It would not only create new job opportunities in the rural regions but would also bring down the debts related to India's energy trade and build up rural financing. Bidding must be done to buy oilseeds from the village committees and farmers. This Competition will help farmers sell oilseeds at higher. Care must be taken that the business-market should be remained unaffected by the oilseed business over the cross-state boundaries.

5 Conclusions and Other Implications

With the increase in the demand of conventional diesel and mandated blending requirements set by the government, the demand for Biodiesel is going to increase in the coming years. In this regard, the development of Biofuels sector in a developing country like India seems to be a highly lucrative option. Subsequently, the Government of India's National Policy on Biofuels have made sufficient efforts

in addressing many precarious issues of food security, rural employment, effective use of unfertile lands etc. relating to the growth of Biodiesel sector in the country. However in spite of these efforts, there are several other problems relating to the Biodiesel production which need to be addressed by the different sectors of the Indian society. The economic viability of Jatropha plantations, successful commercialization of Biodiesel, technical constraints relating to the growth and development of the Biodiesel feedstock plantations, lack of usable data pertaining to the development of Jatropha plantations in India etc. are some of the issues which need immediate attention. Also, a proper resonance and coordination should be maintained between the national and different state governments of India for the effective implementation of the different biofuels policies in the country. In this way, a more environmentally sustainable option of Biofuels sector can be developed which would help in addressing the current Energy crisis of the country.

References

1. Raju SS, Parappurathu S, Chand R, Joshi PK, Kumar P, Msangi S (2012) Biofuels in India: potential, policy and emerging paradigms. NCAP, New Delhi
2. Altenburg T, Dietz H, Hahl M, Nikolidakis N, Rosendahl C, Seelige K (2009) Biodiesel in India: value chain organisation and policy options for rural development. Deutsches InstitutfürEntwicklungspolitik, Bonn
3. Kumar S, Chaube A, Jain SK (2012) Critical review of Jatropha biodiesel promotion policies in India. Energy Policy 41:775–781
4. Leduc S, Natarajan K, Dotzauer E, McCallum I, Obersteiner M (2009) Optimizing biodiesel production in India. Appl Energy 86:S125–S131
5. Biswas PK, Pohit S, Kumar R (2010) Biodiesel from Jatropha: can India meet the 20 % blending target? Energy Policy 38:1477–1484
6. Biswas PK, Pohit S (2013) What ails India's biodiesel programme? Energy Policy 52:789–796
7. Achten WMJ, Almeida J, Fobelets V, Bolle E, Mathijs E, Singh VP, Tewari DN, Verchot LV, Muys B (2010) Life cycle assessment of Jatropha biodiesel as transportation fuel in rural India. Appl Energy 87:3652–3660 (Elsevier)
8. Balat M (2011) Potential alternatives to edible oils for biodiesel production–a review of current work. Energy Convers Manag 52:1479–1492
9. Pohit S, Biswas PK, Kumar R, Goswami A (2010) Pricing model for biodiesel feedstock: a case study of Chhattisgarh in India. Energy Policy 38:1477–1484
10. Demirbas A (2009) Progress and recent trends in biodiesel fuels. Energy Convers Manag 50:14–34 (Elsevier)
11. Jain S, Sharma MP (2010) Prospects of biodiesel from Jatropha in India: a review. Renew Sustain Energy Rev 14:763–771
12. Carraretto C, Macor A, Mirandol A, Stoppato A, Tonon S (2004) Biodiesel as alternative fuel: experimental analysis and energetic evaluations. Energy 29:2195–2211
13. Sahoo PK, Das LM (2009) Combustion analysis of Jatropha. Karanja and Pogamia based biodiesel as fuel in a diesel engine. Fuel 88:994–999
14. Nabi MN, Akhter MS, Shahdat MZ (2006) Improvement of engine emissions with conventional diesel fuel and diesel bio diesel blends. Biosource Technol 91:372–378
15. Greeves G, Wang CHT (1981) Origins of diesel particulate mass emission, SAE. doi:10.4271/810260

A Nominal Stress Based Reliability Analysis Method for Dependent Fatigue and Shock Processes

Hongxia Chen and Yunxia Chen

Abstract The product which has fatigue induced failure mechanism always suffers complex fatigue loads accompanied by random shock loads. The existence of shocks can increase the risk of failure because it can not only cause shock damages or shock failure, but also impact the fatigue process. This chapter discussed the dependent relationship between fatigue and shock processes when high cycle fatigue and low-energy shocks are involved. The uncertainty of external loads and internal properties are considered. The reliability of fatigue process is developed with the stochastic fatigue loads and corresponding stochastic fatigue life model based on the impact of shocks. The reliability of fatigue process is also modelled with the assumption of HPP and the effect of fatigue damages. Overall competitive reliability model is proposed with an engineering case of actuator cylinder.

Keywords Nominal stress · Indeterminacy · Dependent · Competitive reliability

1 Introduction

Fatigue is an important failure mechanism which causes damage to many kinds of products. Fatigue damage is usually caused by working stress which is also called as fatigue load during normal operation. Meanwhile, the products also suffer shock loads caused by foreign object or environment inevitably during the process of manufacturing, transportation, maintenance and operation.

Both the fatigue damage and shock damage have been studied in lots of literatures [1–5]. There are a plenty of fatigue theories based on different definitions of fatigue damages, such as fatigue crack, fatigue life, residual strength, residual toughness, $1/N$ (N denotes fatigue life), and so on. The fatigue theory based on

H. Chen (✉) · Y. Chen
School of Reliability and System Engineering, Beijing University of Aeronautics and Astronautics, Beijing, China
e-mail: Chxxm_123@126.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_102

$1/N$ is usually used to analyze the damages caused by variable amplitude of fatigue loads which is called the accumulative fatigue damage theory. As to the constant amplitude of fatigue loads, the nominal stress method based on the $S - N$ curve is the most widely used fatigue analysis method.

There are also many researches on shock damages, and the most common theories based on statistical theories are cumulative model and extreme model [6, 7]. But they cannot be used to analyze the shock failure before specifying the performance variable.

Some researchers have studied the relationship between fatigue damage and shock damage when both fatigue and shock loads are involved. Liu et al. [8] through the alternate compressive fatigue and impact loading test measured the change of elastic modulus of LRBC. The coupling effects of fatigue damage and impact damage were analyzed by defining the damage as cumulative dissipated energy. The analysis shows that the fatigue damage and impact damage are coupled each other, and the impact damage greatly influences the fatigue evolution. Zhu and Xu [9] performed the test to investigate single impact effects with high strain rate on the low cycle fatigue life of 1Cr18Ni9Ti. The result showed that the effects are dependent with a coupling action of the welding residual stress and the impacted plastic-induced mechanism. Many other researchers have studied the residual strength after shocks. Zhu and Xu [9] discussed the relationship between damage area, compressive strength and impact energy after impact, and analyzed the major damage mechanism of damaged composite under fatigue load. The results showed that the impact damage affected the compressive strength and fatigue performance greatly. Deniz et al. [10] studies the effects of the tube diameter and the impact energy level on the impact and compression after impact behaviour. Results indicated that both specimen diameter and impact energy highly affect the impact response and compression-after impact strength of composite tubes. The above researches indicate that, fatigue and shock damage have a coupling relationship, and the effect of shock to fatigue can be reflected in the residual strength. However, little quantitative analysis has been done on this coupling relationship.

In addition, Chen and Chen [11] developed a coupling model for the fatigue and shock damage, but it only gave a general model without any application. What's more, some random factors such as random fatigue life, random shock stress, and so on, have not been taken into consideration. This chapter proposed a specific reliable analysis method based on the nominal stress with random factors.

The organization of this passage is as follows. In Sect. 2 fatigue damage, shock damage and their relationship are analyzed based on the nominal stress method. The reliable life model is developed by considering the randomness of fatigue and shock loads, as well as the fatigue life in Sect. 3. Section 4 provides an application case of actuator cylinder. And conclusions are given in Sect. 5.

2 Fatigue and Shock Analysis Based on Deterministic Method

2.1 Nominal Stress Fatigue Analysis Method

The traditional nominal stress fatigue analysis method is to analyze the nominal stress of the critical location and use the $S-N$ curve to achieve the fatigue life from nominal stress. This method is effective only when the fatigue load belongs to high cycle fatigue (HCF). However, when the $S-N$ curve is not available, the developed nominal stress method is applied in fatigue analysis. The process of the developed nominal stress method which is applied in variable amplitude fatigue problem is as follows with the assumption that the fatigue load spectrum is consisted of m different stress levels.

(1) Analyze the nominal stress spectrum of the critical location of the component, and get the maximum stress $\sigma_{\max}$ and minimum stress $\sigma_{\min}$.
(2) The Goodman equation is used to modify the average nominal stress to get the equivalent stress σ_s.

$$\sigma_s = \frac{\sigma_a}{1-\frac{\sigma_m}{\sigma_b}} \tag{1}$$

where $\sigma_a = \frac{\sigma_{\max}-\sigma_{\min}}{2}$ is the amplitude of stress and $\sigma_m = \frac{\sigma_{\max}+\sigma_{\min}}{2}$ is the average stress. σ_b is the ultimate of strength.
(3) Determine the fatigue limit σ_{-1A} and its corresponding fatigue life N_0. The Basquin equation is utilized to calculate the fatigue life which is corresponding to the nominal stress of each stress level in the form of Eq. (2).

$$N_i = (\frac{\sigma_{is}}{\sigma_{-1A}})^b N_0 \tag{2}$$

where σ_{is} is the equivalent stress of the ith stress level, b is the fatigue performance related constant, and N_i is the fatigue life of the ith level.
(4) Calculate the accumulative damage by accumulative damage theory. The most common accumulative damage theory is the Miner–Palmgren law:

$$D = \sum_{i=1}^{m} \frac{n_i}{N_i} \tag{3}$$

where n_i is the cycles of fatigue loads in the ith level.

2.2 Shock Analysis

As mentioned in introduction, the shock models cannot be used before giving the specific meaning to the shock variable. In the stress-strength theory, the extreme model is used and the condition of shock failure is

$$\sigma_b(t) \leq \sigma_s(t) \tag{4}$$

where σ_b is the ultimate strength, and σ_s is the stress. The shock variable is the amplitude of stress.

2.3 Coupling Relationship

Present researches [2–9] show that, the shock load will decrease the residual strength and residual fatigue life. Tests of material shows that when the times of shock is less than 500–1,000, the damage law is similar with single shock, but when larger than 100,000, the damage law is same as fatigue damage. Chen assumes that before the 1,000th shock, the ultimate strength of material of product will decrease by the effect of each shock, however, after the 1,000th shock, the ultimate strength will stop depredating because the property of shock is similar as fatigue, and the damage caused by shock will accumulates together with fatigue damage. By using the static strength degradation model, the relationship of shock stress and σ_b after ith shocks can be written as

$$\sigma_b^* = \sigma_b(SS) = \sigma_b(0) - \sum_{j=1}^{\min(i,1,000)} qSS_j^p \tag{5}$$

where p and q are material related constant. When i tends to the corresponding life of the shock stress, σ_b^* is much less than $\sigma_b(0)$, that is the left side of Eq. (5) can be ignored, and hence Eq. (5) can be regarded as a $S - N$ curve. The constant p and q can be derived by $S - N$ curve with shocking test.

3 Reliability Model Based on Indeterminacy Analysis

3.1 Random Fatigue Loads Analysis

The external loads measured in real systems are often quite irregular [12]. This means fatigue loads are varying with time which can be considered a random variable. In order to determine the fatigue damage caused by an irregular load, counting methods are generally adopted. Among the counting methods, the

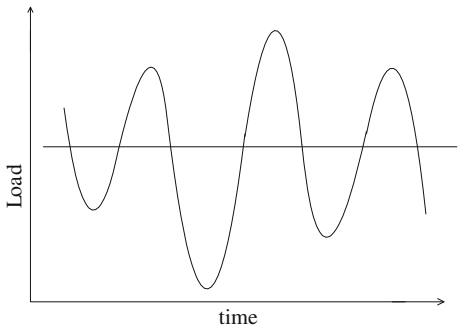


Fig. 1 Random load

rainflow count is widely regarded as the best counting procedure. Hereafter, another counting method is introduced for its simplicity.

As the random load in Fig. 1, the time varying load $x(t)$ can be assumed to only depend on the sequence of extremes (maxima and minima), hence $x(t)$ is given by the sequence $x(t) = \{m_0, M_0, m_1, M_1, m_2, M_2, \ldots\}$. The sequence maybe very long, and the calculation of fatigue damage is complex. An approximate way is to determine a sequence with a fixed level (often 64 or 128, …), and the criterion sequence is

$$\left\{m_0^c, M_0^c, m_1^c, M_1^c, \ldots, m_n^c, M_n^c\right\} \tag{6}$$

We can project the real load sequence to the nearest level, and count the cycles in each level. Then the load spectrum is formed and the random load is changed into multi-amplitude load.

3.2 Random Shock Process

When the arrival of shock load is in random, the Homogeneous Poisson Process (HPP) is used to describe the shock process. Assume the arrival time follows the HPP with parameter λ, and the shock stress SS_i is a random variable follows the same lognormal distribution $LN(\mu_s, \sigma_s^2)$, independently. With the theory in Sect. 2, the ultimate strength σ_b after shock at time t is

$$\begin{aligned}\sigma_b^*(t) &= \sum_{i=0}^{\infty} P\{n(t) = i\}\left(\sigma_b(0) - \sum_{j=0}^{\min(i,1,000)} qSS_j^p\right) \\ &= \sum_{i=0}^{\infty} e^{\lambda t}\frac{(\lambda t)^i}{i!}(\sigma_b(0) - \min(i, 1{,}000)qSS^p)\end{aligned} \tag{7}$$

where $n(t)$ denotes the arrival times of shock process at time t, SS denotes the average shock stress, and $SS = \exp(\mu_s)$.

When $n(t) = i$ and $i \leq 1{,}000$, the equivalent fatigue stress at time t is

$$\sigma_s(t) = \frac{\sigma_a(t)}{1 - \frac{\sigma_m(t)}{\sigma_b^*(t)}} = \frac{\sigma_a(t)}{1 - \frac{\sigma_m(t)}{\sigma_b(0) - \min(i,1{,}000) qSS^p}} \tag{8}$$

The fatigue life between the interval $[t, t+1]$ is

$$N(t) = \left(\frac{\sigma_s(t)}{\sigma_{-1A}}\right)^b N_0 = \left(\frac{\sigma_a(t)/\sigma_{-1A}}{1 - \frac{\sigma_m(t)}{\sigma_b(0) - \min(i,1{,}000) qSS^p}}\right)^b N_0 \tag{9}$$

The fatigue damage caused in this interval is

$$\Delta D(t) = \frac{1}{N(t)} \tag{10}$$

The accumulative fatigue damage caused by

$$D_f(t) = \sum_{\tau=0}^{t} \frac{1}{N(\tau)} = \sum_{\tau=0}^{t} \sum_{i=0}^{\infty} e^{\lambda t} \frac{(\lambda\tau)^i}{i!} \frac{1}{N_0} \frac{\left(1 - \frac{\sigma_m(\tau)}{\sigma_b(0) - \min(i,1{,}000) qSS^p}\right)^b}{\left(\sigma_a(\tau)/\sigma_{-1A}\right)^b} \tag{11}$$

When $n(t) = i > 1{,}000$, the ultimate strength does not degrade with the shock loads, and it keeps at $\sigma_b^*(t) = \sigma_b(0) - \sum_{j=1}^{1{,}000} qSS_j(\tau)^p = \sigma_b(0) - 1{,}000 qSS^p$. The equivalent stress of shock load is

$$\sigma_{ss} = \frac{SS/2}{1 - \frac{SS/2}{\sigma_b(0) - 1{,}000 qSS^p}} \tag{12}$$

The life corresponding to the equivalent stress is

$$N_s = \left(\frac{\sigma_{ss}}{\sigma_{-1A}}\right)^b N_0 = \left(\frac{SS/\sigma_{-1A}}{2 - \frac{SS}{\sigma_b(0) - 1{,}000 qSS^p}}\right)^b N_0 \tag{13}$$

The accumulative fatigue damage caused by shock loads before time t is:

$$D_{sh}(t) = \sum_{\tau=0}^{t} \frac{1}{N_s(\tau)} = \sum_{\tau=0}^{t} \sum_{i=1{,}001}^{\infty} e^{\lambda t} \frac{(\lambda\tau)^i}{i!} \frac{1}{N_0} \frac{\left(2 - \frac{SS}{\sigma_b(0) - 1{,}000 qSS^p}\right)^b}{\left(SS/\sigma_{-1A}\right)^b} \tag{14}$$

Considered the above two kinds of fatigue damage, the total damage at time t can be achieved from Eq. (15):

$$D(t) = D_f(t) + D_{sh}(t) \tag{15}$$

Substitute Eqs. (11) and (14) into Eq. (15), we can get overall fatigue damage. The reliability life can be written as

$$R(t) = R_f(t)R_h(t) \tag{16}$$

where $R_f(t) = P\{D(t) \le D_f\}$ is the probability of no fatigue failure. For the fatigue induced failure, the threshold of failure D_f is usually a probabilistic variable rather than a fixed value because of the dispersion from material or external factors. The specific distribution is decided by material properties and environment, and it can usually be given through engineering experience. Usually, $D_f \sim N(\mu_f, \sigma_f^2)$, then

$$\begin{aligned} R_f(t) &= P\{D(t) \le D_f\} \\ &= 1 - \Phi\left(\frac{D(t) - \mu_f}{\sigma_f}\right) \end{aligned} \tag{17}$$

In Eq. (16), $R_h(t)$ denotes the probability of no shock failure. According to the condition of shock failure determined by Eq. (4), $R_h(t)$ can be derived from the following equations.

The reliability at time $t + \Delta t$ is

$$\begin{aligned} R_h(t+\Delta t) &= P\{\sigma_b^*(\tau) \ge \sigma(\tau) + SS(\tau), \forall \tau \in [0,t]\} P\{\sigma_b^*(\tau) \ge \sigma(\tau) + SS(\tau), \forall \tau \in [t, t+\Delta t]\} \\ &= R_h(t)\left[1 - \lambda \Delta t P\{\sigma_b^*(\tau_0) \le \sigma(\tau_0) + SS, \exists \tau_0 \in [t, t+\Delta t]\} + o(\Delta t)\right] \\ &\quad - R_h(t)(1-\lambda)\Delta t P\{\sigma_b^*(\tau_0) \le \sigma(\tau_0), \exists \tau_0 \in [t, t+\Delta t]\} \end{aligned} \tag{18}$$

As the probability of $P\{\sigma_b^*(\tau_0) \le \sigma(\tau_0), \exists \tau_0 \in [t, t+\Delta t]\} = 0$, when $\Delta t \to 0$, we have $\tau_0 \to t$. Hence,

$$\frac{dR_h(t)}{dt} = -R_h(t)\lambda P\{\sigma_b^*(\tau) \le \sigma(\tau) + SS(\tau)\} \tag{19}$$

The integral of Eq. (19) is

$$\begin{aligned} R_h(t) &= \exp\left(-\lambda \int_0^t P\{\sigma_b^*(\tau) \le \sigma(\tau) + SS\} d\tau\right) \\ &= \exp\left(-\lambda \int_0^t P\left\{\sum_{i=0}^{\infty} e^{\lambda\tau} \frac{(\lambda\tau)^i}{i!} (\sigma_b(0) - \min(i, 1{,}000) qSS^p) \le \sigma(\tau) + SS(\tau)\right\} d\tau\right) \end{aligned} \tag{20}$$

As each of the shock stress SS_i follows the same lognormal distribution $LN(\mu_s, \sigma_s^2)$, independently, then

$$\begin{aligned} &P\left\{\sum_{i=0}^{\infty} e^{\lambda\tau} \frac{(\lambda\tau)^i}{i!} (\sigma_b(0) - \min(i, 1{,}000) qSS^p) \le \sigma(\tau) + SS(\tau)\right\} \\ &= P\left\{\sigma_b(0) - \sum_{i=0}^{\infty} e^{\lambda\tau} \frac{(\lambda\tau)^i}{i!} (\min(i, 1{,}000) qSS^p) - \sigma(\tau) \le SS(\tau)\right\} \end{aligned} \tag{21}$$

Hence

$$\begin{aligned} &P\left\{\sum_{i=0}^{\infty} e^{\lambda\tau} \frac{(\lambda\tau)^i}{i!} \left(\sigma_b(0) - \sum_{j=1}^{\min(i,1,000)} qSS_j^p\right) \le \sigma(\tau) + SS\right\} \\ &= 1 - Ln\Phi\left(\sigma_b(0) - \sum_{i=0}^{\infty} e^{\lambda\tau} \frac{(\lambda\tau)^i}{i!} (\min(i, 1{,}000) qSS^p) - \sigma(\tau), \mu_s, \sigma_s^2\right) \end{aligned} \tag{22}$$

3.3 Random Fatigue Life

On the other hand, if the fatigue life N_0 corresponding to the fatigue limit is considered to follow the lognormal distribution, that is $N_0 \sim LN(\mu_{n0}, \sigma_{n0}^2)$, and then $\frac{1}{N_0} \sim LN(-\mu_{n0}, \sigma_{n0}^2)$.

$$\begin{aligned} D(t) &= \sum_{\tau=0}^{t} \sum_{i=0}^{\infty} e^{\lambda t} \frac{(\lambda\tau)^i}{i!} \frac{1}{N_0} \frac{\left(1 - \frac{\sigma_m(\tau)}{\sigma_b(0) - \min(i,1,000) qSS^p}\right)^b}{(\sigma_a(\tau)/\sigma_{-1A})^b} + \sum_{\tau=0}^{t} \sum_{i=1,001}^{\infty} e^{\lambda t} \frac{(\lambda\tau)^i}{i!} \frac{1}{N_0} \frac{\left(2 - \frac{SS_i}{\sigma_b(0) - 1,000 qSS^p}\right)^b}{(SS_i/\sigma_{-1A})^b} \\ &= \frac{1}{N_0} \left\{ \sum_{\tau=0}^{t} \sum_{i=0}^{\infty} e^{\lambda t} \frac{(\lambda\tau)^i}{i!} \frac{\left(1 - \frac{\sigma_m(\tau)}{\sigma_b(0) - \min(i,1,000) qSS^p}\right)^b}{(\sigma_a(\tau)/\sigma_{-1A})^b} + \sum_{\tau=0}^{t} \sum_{i=1,001}^{\infty} e^{\lambda t} \frac{(\lambda\tau)^i}{i!} \frac{\left(2 - \frac{SS_i}{\sigma_b(0) - 1,000 qSS^p}\right)^b}{(SS_i/\sigma_{-1A})^b} \right\} \end{aligned} \tag{23}$$

Let

$$y(t)=\sum_{\tau=0}^{t}\sum_{i=0}^{\infty}e^{\lambda t}\frac{(\lambda\tau)^i}{i!}\frac{\left(1-\frac{\sigma_m(\tau)}{\sigma_b(0)-\min(i,1{,}000)qSS^p}\right)^b}{\left(\sigma_a(\tau)/\sigma_{-1A}\right)^b}+\sum_{\tau=0}^{t}\sum_{i=1{,}001}^{\infty}e^{\lambda t}\frac{(\lambda\tau)^i}{i!}\frac{\left(2-\frac{SS_i}{\sigma_b(0)-1{,}000qSS^p}\right)^b}{\left(SS_i/\sigma_{-1A}\right)^b}$$

denotes the fatigue coefficient, and we can have $D(t)\sim LN(-y(t)\mu_{n0},y^2(t)\sigma_{n0}^2)$.

Then

$$\begin{aligned}R_f(t)&=P\{D(t)\le D_f\}\\&=Ln\Phi\left(\frac{D_f+y(t)\mu_{n0}}{y(t)\sigma_{n0}}\right)\end{aligned}\tag{24}$$

Through the analysis of Sect. 3.1, the random fatigue loads can be changed into an approximate load spectrum with r stress levels and each has n_i cycles. The stress amplitude $\sigma_a(t)$ and average stress $\sigma_m(t)$ is determined by the load spectrum. The load level $l(t)$ at time t can be obtained from Eq. (25).

$$\begin{aligned}&l(t)=\min l\\&s.t.\\&t-[t/\sum_{i=1}^{r}n_i]\sum_{i=1}^{r}n_i-\sum_{i=1}^{l}n_i\le 0\end{aligned}\tag{25}$$

The symbol $[\cdot]$ means the largest integer in Eq. (25). When $t-[t/L]L=0$, we have $l(t)=r$. For any time t, we have $\sigma_a(t)=\sigma_a(l(t))$ and $\sigma_m(t)=\sigma_m(l(t))$.

4 Case Study

The dominating failure mechanism of an actuator cylinder is the fatigue induced failure. Meanwhile, the actuator cylinder also suffers random shocks. In order to evaluate the fatigue life of the actuator cylinder, the load history is analyzed and the cyclic unit of fatigue load spectrum is listed in Table 1. The original ultimate strength $\sigma_b(0)=1{,}310$ MPa. Besides, the shock process follows a HPP with $\lambda=0.01$ time/cycle, and the shock stress $SS_i\sim LN(\mu_s=6.215,\sigma_s^2=0.0335^2)$. The fatigue load belongs to HCF, and the shock load is in low-energy. Therefore, the method proposed in this chapter can be applied to estimate the reliable life of the actuator cylinder.

The average shock stress $SS=e^{\mu_s}=500$ MPa. The material related constant*s* in Eq. (7) are $p=0.28,q=0.048$. The parameters in Eq. (8) are $b=-3.92$, $\sigma_{-1A}=343$ MPa. If the fatigue life corresponding to the fatigue limit is a fixed value, we also have $N_0=10^7$.

Table 1 Load spectrum of actuator cylinder

Level stage l	Cyclic number n_i	$\sigma_{\max}$ (MPa)	$\sigma_{\min}$ (MPa)	$\sigma_a(l) = \sigma_m(l)$ (MPa)
1	4,000	289	0	144.5
2	8,000	423	0	211.5
3	18,000	534	0	267
4	91,000	423	0	211.5
5	217,000	319	0	159.5
6	1,444,000	301	0	150.5

$$D_f(t) = \sum_{\tau=0}^{t}\sum_{i=0}^{\infty} e^{\lambda\tau}\frac{(\lambda\tau)^i}{i!}\frac{1}{N_0}\frac{\left(1-\frac{\sigma_m(\tau)}{\sigma_b(0)-\min(i,1{,}000)qSS^p}\right)^b}{\left(\sigma_a(\tau)/\sigma_{-1A}\right)^b}$$
$$= \sum_{\tau=0}^{t}\sum_{i=0}^{\infty} e^{-\lambda\tau}\frac{(\lambda\tau)^i}{i!}\frac{1}{10^7}\frac{\left(1-\frac{\sigma_m(l(\tau))}{1{,}310-0.48\times 500^{0.036}\min(i,1{,}000)}\right)^{-3.92}}{\left(\sigma_a(l(\tau))/343\right)^{-3.92}}$$

and

$$D_{sh}(t) = \sum_{\tau=0}^{t}\sum_{i=1{,}001}^{\infty} e^{\lambda\tau}\frac{(\lambda\tau)^i}{i!}\frac{1}{N_0}\frac{\left(2-\frac{SS}{\sigma_b(0)-1{,}000qSS^p}\right)^b}{\left(SS/\sigma_{-1A}\right)^b}$$
$$= \sum_{\tau=0}^{t}\sum_{i=1{,}001}^{\infty} e^{\lambda\tau}\frac{(\lambda\tau)^i}{i!}\frac{1}{10^7}\frac{\left(2-\frac{500}{1{,}310-1{,}000\times 0.48\times 500^{0.036}}\right)^{-3.92}}{(500/343)^{-3.92}}$$

Equation (16) is used to obtain the probability of no fatigue failure when the shock is random, and the reliability curve is shown in Fig. 2 which is plotted by red

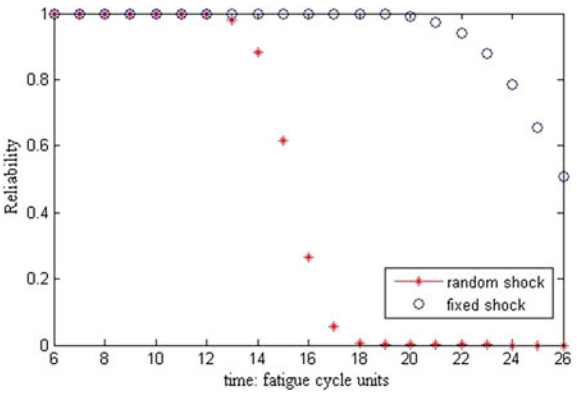


Fig. 2 Reliability curve of no fatigue failure considering random shock and fixed shock

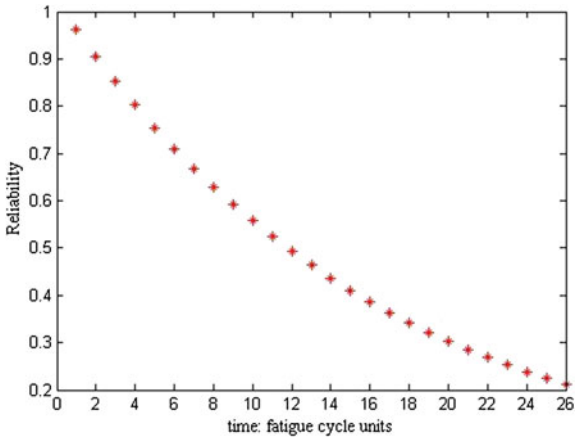


Fig. 3 The probability of no shock failure

asterisk. In addition, the probability of no fatigue failure when the shock arrives fixed is also shown in Fig. 2. We can find that the product fails at 15 units for the largest probability when the shock is random, while the product fails at 26 units for the largest probability when the shock is fixed.

As to probability of no shock failure, $SS \sim LN(6.215, 0.0335^2)$, then we have

$$\begin{aligned} R_h(t) &= \exp\left\{-\lambda t + \lambda Ln\Phi\left(\sigma_b(0) - \sum_{i=0}^{\infty} e^{\lambda\tau}\frac{(\lambda\tau)^i}{i!}(\min(i,1{,}000)qSS^p) - \sigma(\tau), \mu_s, \sigma_s^2\right)d\tau\right\} \\ &\approx 1 \times \exp\left\{-\lambda t + \lambda \int_{120{,}000}^{t} Ln\Phi\left(\sigma_b(0) - 1{,}000qSS^p - \sigma(\tau), \mu_s, \sigma_s^2\right)d\tau\right\} \\ &= \exp\left\{-\lambda t + \lambda \int_{120{,}000}^{t} Ln\Phi\left(1{,}036.5 - \sigma(\tau), 6.215, 0.0335^2\right)d\tau\right\} \end{aligned}$$

The probability of no shock failure is plotted in Fig. 3. As we can see, the product fails at 10th units for the largest probability. Combined with the fatigue failure, the product fails at the 10th units for the largest probability.

5 Conclusions

This chapter proposes an approach to estimate the reliability life for the product suffers fatigue induced failure based on both fatigue load and random shocks. Based on the assumption of HCF and low-energy shocks, the nominal stress method was used to analyze the fatigue damage and shock damage. The coupling relationship between them is reflected on the degradation of ultimate strength of material

directly, and therefore the damage caused by fatigue load after each shock increases within the 1,000th shock. Reliable life models are developed when the random factors are considered, which include random fatigue loads, random shock process and the dispersion of fatigue life. Finally, an engineering case utilizes the proposed approach to estimate reliability life of actuator cylinder which shows the applicable of this approach. The result shows that the risk of estimation is much higher when the random factors are not considered. It provides a more precise way to predict the life of mechanical and electrical product as a basis of reliability related test design, making a more reasonable maintenance schedule.

References

1. Paris PC, Erdogan FA (1963) Critical analysis of crack propagation laws. J Basic Eng ASME (Series D) 85:528–534
2. Shokrieh MM, Lessard LB (1997) Multiraxial fatigue behaviour of unidirectional plies based on uniaxial fatigue experiments-part I. Modelling. Int J Fatigue 19(3):201–207
3. Shen S, Carpenter SH (1929) Application of the dissipated energy concept in fatigue endurance limit testing. J Transp Res Board 2005:165–173
4. Coffin LF (1954) A study of the effects of cyclic thermal stresses on a ductile metal. Trans Am Soc Mech Eng 76:931–950
5. Hwang W, Han KS (1986) Cumulative damage models and multi-stress fatigue life prediction. J Compos Mater 20:125–153
6. Esary J, Marshall A, Proschan F (1973) Shock models and wear process. Ann Probab 1 (17):627–649
7. Fan J, Ghurke SG, Levine RA (2000) Multi-component 1ifetime distribution in the presence of ageing. J Appl Probab 37:521–533
8. Liu YP, He TH, Huang XQ, Tang LQ (2010) Coupling analysis on impact damage and fatigue damage based on dissipated energy theories. Acta Armamentarii 31(1):223–226
9. Zhu WY, Xu XW (2012) Experiment research on residual compressive strength and fatigue performance of composite laminates with low velocity impact damage. Acta Materiae Compositae Snica 29(5):172–178
10. Deniz ME, Karakuzu R, Sari M, Icten BM (2012) An approach for prediction of fatigue life based on fatigue and low-energy shocks. J Compos Mater 46(6):737–745
11. Chen H, Chen Y (2013) An approach for prediction of fatigue life based on fatigue and low-energy shocks. In: Proceeding of QR2MSE
12. Benascuitti D (2004) Fatigue analysis of random loading. University of Ferrara, Italy

Study of Li-Ion Cells Accelerated Test Based on Degradation Path

YunLong Huang and XiaoGang Li

Abstract According to the characteristics of long life, high reliability of Li-ion cells, a constant stress accelerated degradation test method is put forward to assessing reliability and predict lifetime of Li-ion cells. First, FMEA analysis is carried out on Li-ion cells, and the most common failure mode is capacity reduction which the temperature is the most important accelerated stress. Second, the accelerated degradation model and the parameters degradation path model of Li-ion cells are determined, and the accelerated degradation test of Li-ion cells is designed through the analysis of sensitive stresses. With the method of Bartlett statistics, the degradation mechanism consistency boundary is determined on the base of parameter degeneration path in order to obtain the accelerated stress level. The reliability of Li-ion cells is assessed based on pseudo lifetime.

Keywords Li-ion cells · FMEA · Capacity reduction · Accelerated degradation test · Bartlett statistic

1 Introduction

Li-ion cells are independent power supplies of good performance and long life. Since the superiority of Li-ion cells, its breadth of application becomes wider and wider, including the automobile traffic, the satellite spacecraft, the backup power supply and the energy reserves. Li-ion cells play an increasingly important role. Therefore, the requirements of its performance, lifetime and reliability are also increasing.

Y. Huang (✉) · X. Li
School of Reliability and System Engineering, Beihang University,
Beijing, China
e-mail: huangyunlong3108@163.com

X. Li
e-mail: lxg@buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_103

The products of Li-ion cells not only have long-life and high reliability features, but also update quickly. To obtain the credible lifetime and reliability targets of Li-ion cells, in the case of saving time and cost, the appropriate testing techniques and life assessment methods must be used. If the accelerated life testing is used for assessing the reliability of products, a sufficient number of failure data must be required. Obviously, this is inappropriate for the products of long life and high reliability. Accelerated degradation testing is a method of accelerating the degradation performance of products by increasing the levels of stress and collecting data of performance degradation in the condition of the same degradation mechanism. The degradation data are used to assess the reliability of products and extrapolate the lifetime of highly reliable products under normal use conditions statistically. Considering the implementation of constant stress accelerated degradation testing is simple and processing data easily, constant stress accelerated degradation testing for Li-ion cells are studied in this chapter. First, FMEA analysis is carried out on Li-ion cells. Using accelerated degradation testing, the degradation mechanism consistency boundary is determined with the method of Bartlett statistics. By the integration of the accelerated model and the degradation model of performance parameters, the reliability and lifetime of Li-ion cells are assessed based on pseudo lifetime.

2 Weak Link Analysis and Sensitive Stresses Analysis of Li-Ion Cells

2.1 FMEA of Li-Ion Cells

After reading a lot of research articles about the failure mechanism of Li-ion cells [1–4], the common failure modes of Li-ion cells include capacity reduction, liquid leakage, gas leakage and thermal runaway etc. FMEA analysis is carried out on Li-ion cells according to its failure mechanism and failure modes. The causes of degradation are summarized in the order of "failure cause—effects—performance degradation—sensitive stresses" in Table 1.

The analysis shows that failure mechanisms of Li-ion cells are complicated, but the most common failure mode of Li-ion cells is capacity reduction. Li-ion cells can identify capacity reduction failure when the capacity of Li-ion cells decays to below 70 % of the initial capacity [5]. In the design of accelerated degradation test, the capacity of Li-ion cells can be used as performance degradation parameter.

2.2 Sensitive Stresses of Li-Ion Cells

Capacity reduction is caused by various failure mechanisms for Li-ion cells. Temperature and charging current are two main stresses of accelerating the capacity reduction of Li-ion cells. Since the charging system of Li-ion cells is fixed, the case

Table 1 Failure mode and effect analysis of Li-ion cells

Failure causes	Effects	Degradation performance	Sensitive stresses
Electrolyte decomposition (slow boundary effect)	Lithium loss, resistance increases	Energy capacity	High temperature high SOC
Solvent gas precipitation, molecules cracking	Lithium loss, active substance loss	Capacity	Overcharge
The reduction of contact areas as a result of the growing passivation layer	Resistance increases	Energy	High temperature high SOC overcharge
Porosity changes	Resistance increases	Energy	High-frequency cycles high SOC
The active substance loss caused by voltage's changes	Active substance loss	Capacity	High-frequency cycles high DOD overcharge
Adhesive decomposition	Lithium loss, decreased stability	Capacity	High temperature high SOC
Collector corrosion	Resistance increases, disproportionation	Energy	Over-discharge low SOC
Lithium precipitation	Li-ion loss	Capacity	High-frequency cycles hypothermia
Volume change of lattice with deintercalation	Electrode phase transition	Energy Capacity	High temperature high SOC

Note SOC state of charge, 0–100 %, the percentage of actual storage capacity; *DOD* depth of discharge, 0–100 %, the percentage of electricity discharge

of charging current is generally constant. Thus, the only stress used in the testing is temperature. Figure 1 shows the influence of different temperatures on cycle life, and the capacity of Li-ion cells reduces faster in the higher temperature.

3 Li-Ion Cells Failure Models Based on Accelerated Degradation

3.1 Accelerated Degradation Models

According to accelerated degradation models, the relationship between temperature and lifetime of Li-ion cells is established. Furthermore, the lifetime of Li-ion cells

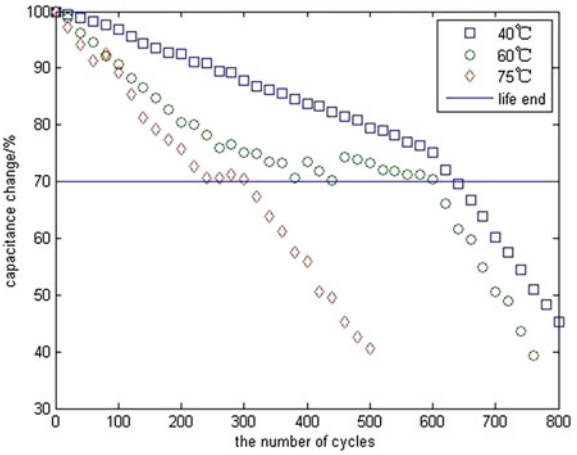


Fig. 1 Influence of temperature on cycle life

under normal use conditions can be extrapolated. Before accelerated degradation models are determined, the assumption is put forward: Under different accelerated stress levels S_i, pseudo lifetime distribution of Li-ion cells obeys Weibull distribution $W(m_i, \eta_i), i = 0, 1, \ldots, k$.

In accelerated degradation testing, temperature is used as the accelerated stress. Therefore, the Arrhenius model is chosen. Because it is assumed pseudo lifetime distribution of Li-ion cells is Weibull distribution, characteristic life η is used as the indicator. For convenience, logarithmic transform of the Arrhenius model is made:

$$\ln \eta = a + b/T \tag{1}$$

In this equation, η represents the characteristic life of Li-ion cells. It is $a = \ln A$, and A is constant. With $b = E/K$, E is activation energy and K is Boltzman's constant. T is thermodynamic temperature stress.

3.2 Accelerated Degradation Path Model

For Li-ion cells, the lifetime is charge–discharge cycles when the capacity of Li-ion cells decays to below 70 % of the initial capacity. In the constant stress accelerated degradation test, there are n_i samples in every stress S_i. The performance degradation data of different testing samples, (t_j, $y_{ik}(t_j)$), are measured in test periods $t_1, t_2, \ldots, t_{mi}$ (i = 1, 2, …, d; j = 1, 2,…, m_i; k = 1, 2,…, n_i). The curve that capacity data for each sample is changed with the charging and discharging cycles is drawn under different stress levels S_i. The degradation curve of Li-ion cells capacity data is fitted with the appropriate degradation paths. When capacity reduction data of Li-ion cells are measured, it generally requires that capacity data are recorded in each cycle. However, if the number of cycles is very large, capacity data can be measured once each certain number of cycles.

The performance degradation of Li-ion cells is observed according to the cumulative damage principle, and the capacity reduction data of Li-ion cells will monotonically decrease with time variation. The capacity data of the kth sample under the S_ith stress level should meet: $y_{i1k} \leq y_{i2k} \leq \cdots \leq y_{ijk} \leq \cdots \leq y_{im_ik}$. Therefore, the degradation path model of Li-ion cells is built:

$$y_{ijk} = g(t_j, \beta_i) + \varepsilon_{ijk} \tag{2}$$

where y_{ijk} is the capacity measurement of the kth sample with t_i charge–discharge cycles under the S_ith stress level, β_i is the parameter vector related with accelerated stress, and ε_{ijk} is the measurement error that obeys the normal distributions with zero mean.

The parameters of the degradation path model can be estimated by the least squares method. So the degradation path of each Li-ion cell can be obtained under each stress level. Then the pseudo lifetime of each Li-ion cell is determined according to the failure threshold value.

4 Accelerated Degradation Testing Design of Li-Ion Cells

4.1 Consistency Test of Failure Mechanisms

Before accelerated degradation testing, the accelerated stress levels must be determined accurately. Accelerated degradation test of Li-ion cells is conducted on the premise that the degradation mechanisms remain consistent under high stress levels. It means that improving the stress level only increases degradation rate without changing the degradation mechanisms [6, 7]. If the stress level isn't high enough, it would not achieve the effect of accelerated degradation. So determining accelerated stress level is very important. To improve the accuracy of Li-ion cells accelerated degradation test and determine the degradation mechanism consistency boundary, the method of Bartlett statistics is presented.

Under each temperature stress level S_i, the pseudo lifetime $t_{ik}(i = 1, 2, \ldots, d; k = 1, 2, \ldots, n_i)$ of each Li-ion cell obeys Weibull distribution. Take the logarithm of t_{ik}, namely

$$x_{ik} = \ln t_{ik}(i = 1, 2, \ldots, d; k = 1, 2, \ldots, n_i). \tag{3}$$

x_{ik} obeys extreme value distribution $G(\mu_i, \sigma_i)$, where it is $\mu_i = \ln \eta_i, \sigma_i = 1/m_i$. x_{ik} is ordered from small to large, and $\hat{\sigma}$ is obtained according to best linear unbiased estimate (BLUE). The variance of $\hat{\sigma}$ is

$$Var(\hat{\sigma}_i) = l_{n_i n_i}\sigma^2, i = 1, 2, \ldots, d \tag{4}$$

In the Eq. (4), $l_{n_i n_i}$ is variance coefficient of $\hat{\sigma}_i$ and it can be acquired through Hand-book of table for reliability testing [8]. $2l_{n_i n_i}^{-1}\hat{\sigma}_i/\sigma_i$ approximately obeys χ^2 distribution with degree of freedom $2l_{n_i n_i}^{-1}$.

Determine degradation mechanism consistency boundary, namely demonstrate the shape parameter consistency under different temperature stress levels $S_1, S_2, \ldots, S_i$ [9]:

$$\mathrm{H}_0 : \mathrm{m}_1 = \mathrm{m}_2 = \cdots = \mathrm{m}_\mathrm{d}$$

Because it is $\sigma_i = 1/m_i$, it is equivalent to demonstrate that the various σ_i are equal:

$$\mathrm{H}_0' : \sigma 1 = \sigma 2 = \cdots = \sigma \mathrm{d}$$

According to Bartlett statistics [10], it shows:

$$\mathrm{B}^2 = 2\left(\sum_{\mathrm{i}=1}^{\mathrm{d}} \mathrm{l}_{\mathrm{n_i n_i}}^{-1}\right)\left[\ln\left(\sum_{\mathrm{i}=1}^{\mathrm{d}} \mathrm{l}_{\mathrm{n_i n_i}}^{-1}\hat{\sigma}_\mathrm{i}\right) - \ln\left(\sum_{\mathrm{i}=1}^{\mathrm{d}} \mathrm{l}_{\mathrm{n_i n_i}}^{-1}\right)\right] - 2\sum_{\mathrm{i}=1}^{\mathrm{d}} \mathrm{l}_{\mathrm{n_i n_i}}^{-1} \ln \hat{\sigma}_\mathrm{i} \tag{5}$$

$$\mathrm{C} = 1 + \frac{1}{6(\mathrm{d}-1)}\left[\sum_{\mathrm{i}=1}^{\mathrm{d}} \mathrm{l}_{\mathrm{n_i n_i}} - \left(\sum_{\mathrm{i}=1}^{\mathrm{d}} \mathrm{l}_{\mathrm{n_i n_i}}^{-1}\right)^{-1}\right] \tag{6}$$

If H_0' is available, B^2/C approximately obeys χ^2 distribution with degree of freedom d − 1 with the specified significance level α. When it is $\mathrm{B}^2/\mathrm{C} > \chi_\alpha^2(\mathrm{d}-1)$, H_0' is refused. When boundary stress is determined with Bartlett statistics, it should be demonstrated one by one, namely demonstrating whether the first two stresses σ_1 and σ_2 are equal. If it is $\sigma_1 = \sigma_2$, demonstrate the third stress σ_3, testing $\sigma_1 = \sigma_2 = \sigma_3$. This continues until σ_d is tested under boundary stress condition, thereby degradation mechanism consistency boundary is determined.

4.2 Accelerated Degradation Testing Design

The operating temperature range of Li-ion cells is 0–40 °C [11]. The higher the temperature is, the faster the capacity of Li-ion cells reduces. Therefore, the temperature stress level of failure mechanism consistency must be determined before accelerated degradation testing. According to Bartlett statistics, significance level is set as 0.05, using the Eqs. (5) and (6), the boundary temperature stress of Li-ion cells is determined. Three different temperature stresses are chosen as accelerated stresses, namely 45, 60, 75 °C. After calculating, the failure mechanism of Li-ion

cells isn't changed under these 3 different temperature stress levels. The reference [12] shows that Li-ion cells will not explode, leak and appear other unexpected failures as long as the temperature does not exceed 90 °C.

Five Li-ion cells of the same initial capacity, model and batch are chosen under each temperature stress level. A standard method for charging and discharging is used. The charging current (power factor) is 1 C, the charging termination voltage is 4.2 V, and the discharging termination voltage is 3 V. It cycles 200 times like this, and capacity data for each sample are recorded once each ten cycles. Capacity data for each sample are $y_{ik}(t_j), i = 1, 2, 3; k = 1, \ldots, 5; t_j = 1, \ldots, 20$. According to degradation data, the degradation path fitting for every product is conducted under each stress level. Combined with the failure criterion of Li-ion cells capacity reduction, the pseudo lifetime of every sample is t_{ik} under every stress level. The pseudo lifetime t_{ik} is the charging–discharging cycle number that the capacity of Li-ion cells reduces to a predetermined threshold.

5 Reliability Assessment of Li-Ion Cells Based on Degradation Data

Reliability assessment of Li-ion cells based on degradation data is to estimate the reliability indicators in the normal stress level through the life characteristics of products in the high stress levels, combined with the accelerated degradation equations. Firstly according to the failure threshold, the pseudo lifetime of every Li-ion cell t_{ik} is obtained in each stress level S_i. According to the order from small to large, the pseudo lifetimes are arranged as $t_{i1} \le t_{i2} \le \cdots \le t_{ij} \le \cdots \le t_{in}$, taking their logarithm as $x_{i1} \le x_{i2} \le \cdots \le x_{ij} \le \cdots \le x_{in}$. Secondly using BLUE methodology, parameter estimation $\hat{\mu}_i$ of extreme value distribution μ_i is calculated. Finally, $\mu_i = \ln \eta_i$, so there is $\mu_i = \ln \eta_i$. The accelerated equation in each stress level is obtained:

$$\ln \eta_i = a + b/T_i \tag{7}$$

According to d different stress levels, characteristic lifetimes of d different Li-ion cells under corresponding temperature stress levels are achieved, namely $(T_i, \hat{\eta}_i), i = 1, 2, \ldots, d$. $\hat{a}, \hat{b}$ are obtained according to the least squares estimation. So the life distribution parameter of Li-ion cells in normal stress level is acquired:

$$\hat{\eta_0} = \exp(\hat{a} + \hat{b}/T_0) \tag{8}$$

Because degradation mechanism maintains consistent in accelerated degradation test of Li-ion cells, shape parameter m_0 maintains constant in normal stress level T_0

and accelerated stress levels $T_i(i = 1, 2, \ldots, d)$. Therefore, the reliability of Li-ion cells in the normal operating temperature is obtained with specified Charging and discharging cycles:

$$\hat{R}(t) = \exp[-(\frac{t}{\hat{\eta}_0})^{m_0}] \tag{9}$$

6 Conclusion

According to FMEA analysis, the most common failure mode of Li-ion cells is capacity reduction. Temperature and charging current are two main stresses of accelerating capacity reduction. In the accelerated degradation testing of Li-ion cells, the degradation mechanism consistency boundary is determined with the method of Bartlett statistics. This method can effectively determine the boundary stress of Li-ion cells degradation testing and improve the accuracy and reliability of the test. Finally, the reliability of Li-ion cells is assessed based on pseudo lifetime.

References

1. Jia Y, Li HL (2008) Research on failure rate model of lithium—ion batteries. J Beijing Univ Aeronaut Astronaut 34(8):974–975
2. Thomas EV, Case HL, Doughty DH (2003) Accelerated power degradation of Li-ion cells. J Power Sources 123(7):254–260
3. Li HL, Su JR (2008) Cycle–life prediction model studies of lithium-ion batteries. Chin J Power Sources 1(4):242–246
4. Huang KL, Lv ZZ, Liu SQ (2001) On capacity fading and its mechanisms for lithium-ion batteries. Battery Bimonthly 31(3):142–144
5. Asakura K, Shimomura M, Shoda T (2003) Study of life evaluation methods for Li-ion batteries for backup applications. J Power Sources 119(121):902–905
6. Feng J (2011) Consistent test of accelerated storage degradation failure mechanism based on rank correlation coefficient. J Aerosp Power 26(11):2440–2444
7. Hu JM, Barker D, Dasgupsta A (1993) Role of failure-mechanism identification in accelerated testing. J IES 26(4):39–45
8. China Electronics Technology Standardization Institute (1987) Hand-book of table for reliability testing. National Defence Industry Press, Beijing
9. Lin FC, Wang QC, Chen YX (2012) Pseudo-life-based test method of mechanism consistency boundary for accelerated degradation testing. J Beijing Univ Aeronaut Astronaut 38(2) (in Chinese)
10. Bartlett MS (1937) Properties of sufficiency and statistical test. In: Proceedings of the Royal Society
11. Wu Y, Jiang XH, Xie JY (2009) The reasons of rapid decline in cycle life of Li-ion battery. Battery Bimonthly 39(4):206–207
12. Xie JY (2005) Study on safety of lithium-ion batteries. Institute of microsystem and information technology, Shanghai (CAS, 70–79)

Applicability Study on Fault Diagnostic Methods for Analog Electronic Systems

Rongbin Guo, Shunong Zhang, Peng Gao and Jiaming Liu

Abstract In this article, a comprehensive investigation to the current fault diagnostic methods for analog circuits is conducted. The ideas of these methods are described. Some key capabilities of these methods for analog circuits are taken into account, which include the capabilities of detection, location and identification, the capabilities of detecting single fault or multiple faults and soft faults or hard faults, the capabilities of diagnosing linear circuits or nonlinear circuits, etc. Then, what kinds of methods are applied to the built-in test (BIT) for electronic systems is investigated. At last, summary and future work on the research are presented.

1 Introduction

Prognostics and Health Management (PHM) is the key technology to achieve Condition Based Maintenance (CBM) and Automatic logistic (AL), where PHM applied to Electronic Systems is one of the most important application areas, while fault diagnosis is a prerequisite for fault prognosis. Electronic systems are generally composed of digital and analog circuits. Since the 1960s, a series of exploration on the topics of fault diagnosis are conducted. So far, some satisfactory results have been made in the digital circuit fault diagnosis; however, the development of fault diag-

R. Guo
Science and Technology on Electronic Test and Measurement Laboratory, Qingdao, China
e-mail: eiqd@ei41.com

S. Zhang (✉) · P. Gao · J. Liu
National Laboratory for Reliability and Environmental Engineering, School of Reliability and Systems Engineering, Beihang University, Beijing, China
e-mail: zsn@buaa.edu.cn

P. Gao
e-mail: 453632324@qq.com

J. Liu
e-mail: liujiaming@dse.buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_104

nosis on analog circuits has been slow. The main reason is that the fault diagnosis on analog circuit encountered some difficulties that the fault diagnosis on digital circuit does not have, which mainly come from the analog component tolerances and nonlinear properties. According to the literature [1], although 80 % of the part of electronic system is composed of digital circuits, more than 80 % of the faults come from the analog circuits. Therefore, the fault diagnosis on analog circuit has been a "bottleneck" problem on the development of electronic industry for a long time.

The fault diagnostic techniques on analog circuits has been reviewed by some literatures, such as Duhames [2–4], etc. This chapter just focuses on the applicability of the methods of fault diagnosis for analog electronic systems.

2 Classification of the Fault Diagnostic Methods for Analog Circuits

Based on the circuit simulation is before or after test, the fault diagnostic methods for analog circuits can be divided into: simulation before test and simulation after test, as shown in Fig. 1.

The diagnostic methods of simulation before test mainly include the fault dictionary method and probability method, and the most basic and typical method is the fault dictionary method. The diagnostic methods of simulation after test mainly include parameter identification method, network decomposition method, fault verification method and artificial intelligence method which developed rapidly in recent years.

In order to be able to monitor real-time electronic system health condition, the software with online diagnosis and prognosis is needed. Furthermore, we need to know which method can be used for fault detection, which method can be used for fault location, and which method can be used for fault identification. Also, we need to know under which conditions the fault diagnostic methods can quickly and accurately diagnose faults, namely the selection of best method.

To conduct PHM, method selection needs to be finished at the design stage of products, therefore, the applicability of the methods for fault diagnosis mentioned above, which mainly include the following aspects, need to be clearly indicated.

1. Whether the method can finish the three processes of fault diagnosis, namely the detection, isolation and identification. Some methods can isolate the fault to the system level, some can isolate to the subsystems or module levels, and some can isolate to single component. There are also some methods which can identify the degrees of faults.
2. Whether the method can be applicable to linear circuits or nonlinear circuits. A linear circuit is the circuit composed entirely of linear elements, such as the circuit composed entirely of resistance, power source and other linear elements. A nonlinear circuit is the circuit containing any nonlinear elements. By means of the circuit analysis for linear circuits, we can obtain the circuit equations, while it is difficult to obtain the circuit equations of nonlinear circuits.

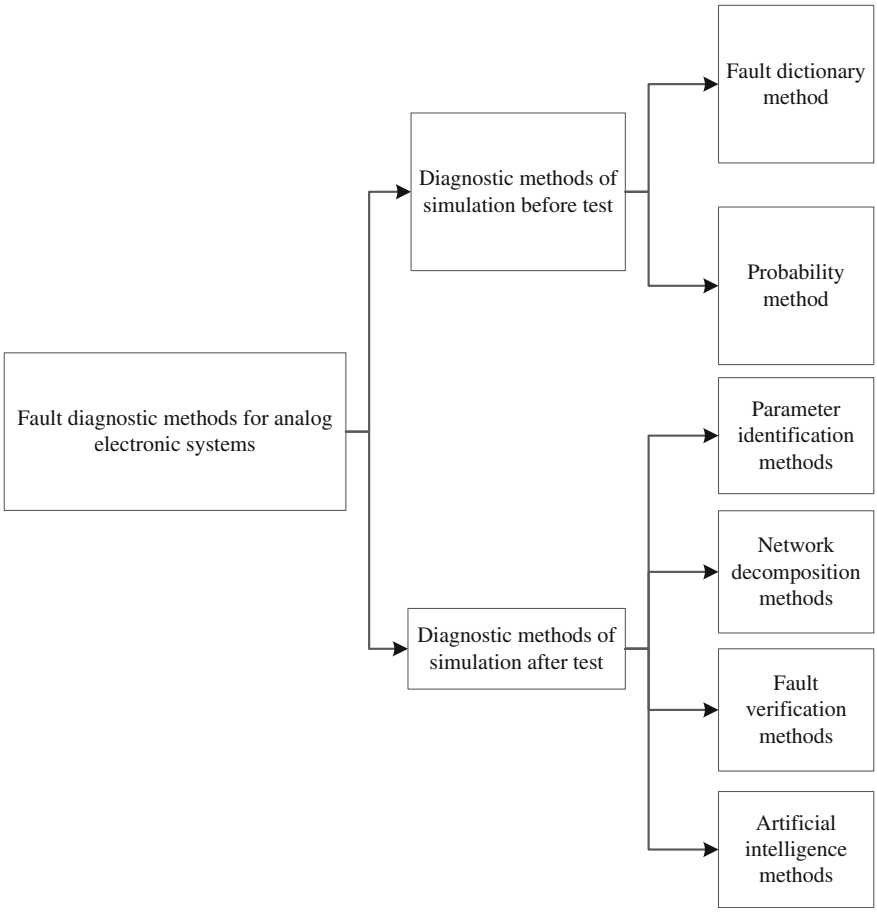


Fig. 1 Classification of fault diagnostic methods for analog electronic systems

3. Whether the method can be used to diagnose single fault or multi faults. Some fault diagnostic methods are based on a single fault assumption, namely there is only one element fails at one time. But in practical situations, multiple faults may appear. So the diagnostic methods we select should correspond to the actual diagnostic requirements.
4. Whether the method can be used to diagnose hard faults or soft faults. The faults of circuit elements can be divided into hard and soft faults. Hard faults refer to some outburst faults like open or short circuit faults of the elements. Soft faults refer to the parameters of these elements exceeding their tolerances which mean the elements do not completely fail. Generally, the methods used for soft fault diagnosis can diagnose hard faults, because a hard fault can be regarded as a special case of a soft fault.

3 The Characters and Applicability of the Fault Diagnostic Methods for Analog Circuits

3.1 Diagnostic Methods of Simulation Before Test

The diagnostic methods of simulation before test mainly include fault dictionary method and probability method:

1. The idea of fault dictionary method [5]: Circuit simulations need to be conducted firstly to find fault characteristics. Then a fault dictionary needs to constitute according to the fault characteristics of almost all kinds of faults. The measured values of characteristics of the circuits under test need to be compared with the values in the fault dictionary so as to identify the circuit faults when it was used for diagnosis.
2. The idea of probability method [6]: According to the measured data of system parameters and their distribution, statistical techniques are used to determine the fault probabilities of the elements with the related circuit parameters. The probability methods include inverse probability method, statistical simulation method and statistical inference method.

3.2 Diagnostic Methods of Simulation After Test

The diagnostic methods of simulation after test mentioned will be described respectively below.

3.2.1 Parameter Identification Methods

Parameter identification methods [5] are those methods whose purpose is to identify all the parameters of the circuit network through testing in order to determine the fault locations. Using the analytic relationship between the circuit network characteristics and the component parameters, and calculating all the element parameter values through the measurement values, fault elements can be determined according to whether these values are within the tolerance ranges or not. The methods are especially suitable for detection of soft faults. Based on the characteristics of diagnostic equations, the method can be divided into two kinds: linear and nonlinear. Generally, the global unique solution usually can be gotten for linear diagnostic equations, while the local unique solution can be gotten for nonlinear diagnostic equations [7]. The parameter identification methods mainly include parameter estimation methods, Navid transfer admittance parameter method, adjoint circuit method, component connection model method, equivalent circuit conversion method, etc.

1. Parameter estimation methods [5] are approximate methods which in most cases require only a small amount of measurement data to estimate the most likely failure element by using the estimation techniques. The parameter estimation methods include combined criteria method, iterative method, quadratic programming method, L1 norm method, symbolic transfer function method, voting method, minimum standard deviation method, potential function method, impact function method, etc.

   - Combined criteria method [5]: For this method, the most probable fault parameters can be determined by finding the parameters corresponding to the smallest characteristic deviation. The advantage of this method is that the few measurement data are needed.
   - Iterative method [5, 8]: For this method, a criteria which is similar to a sensitivity factor is used, and the value of each parameter when the criteria reaches the minimum, which is the actual value of each parameter, is calculated and then is compared with its rated value to determine the fault location.
   - Quadratic programming method [5]: This method is based on the least squares method. The Merrill optimization method and linearization process are used to establish standard quadratic programming problems, so as to achieve the goal of fault location and identification.
   - L1 norm method [5, 7]: The idea of this method is to establish the optimized mathematical model by the character that the sum of residual squares of first order norm is equal to the sum of the absolute value of each component. Since the objective function and constraints are linear, it belongs to linear programming method. The deviations of all the components need to be calculated under the constraints of equations and the minimum of the objective function. Then, any element, whose deviation is beyond the tolerance limit, can be diagnosed as fault element.
   - Symbolic transfer function method [5]: For this method, the symbolic transfer function formula of the circuits to be diagnosed needs to be calculated firstly, then a set of circuit parameters which satisfy the frequency response of the circuit which is most close to that of the circuit to be diagnosed can be found by optimization methods. Thus the values of the set of circuit parameters are the approximate value of the parameters of the fault circuit, and then the fault components can be isolated. Due to the adoption of conjugate direction method to optimize the symbolic transfer function method, the convergence speed is quickly.
   - Voting method [5]: Its idea is as follows: Compare the characteristics of bias and sensitivity matrix to identify the fault component; for the elements of sensitivity matrix, only its symbol needs to be considered, and its values can be ignored. The advantage of this method is that the diagnose process is very simple and the probable fault sequence can be arranged by diagnostic results.

- The minimum standard deviation method [5]: After excitation is implement to the diagnostic circuit, the measured characteristic values (assume there are m values) can be gotten, and they can be compared with the related normal characteristic values, and then the deviations of the m characteristic values can be gotten. Assume that these characteristics deviations are generated by a parameter deviation of a fault component. According to the deviation of each characteristic value and the sensitivities that the characteristic value is related to this parameter, a value for the parameter deviation can be calculated. So m values can be gotten, and then the fault component can be diagnosed by the distributions of the m values.
- Potential function method [5]: Use the potential function, under a certain recursive procedure, to get the deviation of response of the circuit to be diagnosed. This method is effective when a certain parameter deviation is much greater than others. This method can make full use of the large amount of data recorded in normal operation, and can save the computer capacity. However, this method should not be used when the measured data are not all ready.
- Impulse function method [5]: This method is based on the impulse responses of the circuit under test to estimate circuit parameters. Under the condition that the circuit parameters are all rated value, the sensitivity of the impulse responses of each discrete point to the circuit parameters are considered. Then the deviations of impulse responses between measured values and rated values under normal conditions are calculated to determine the fault components.

2. Navid transfer admittance method [5–7]: For this method, the nonlinear equations of the transfer admittance parameters of circuit network ports need to be solved to determine the network element values. Its advantage is ease to test, while its shortcoming is that the transfer admittance function is nonlinear. Therefore a logarithmic transformation is needed to convert it to quasilinear problems.
3. Adjoint circuit method [5, 9–12]: For this method, the Tellegen's theorem and the adjoint circuit are used to establish fault circuit equations, so as to calculate incensement of the parameters of each element in the circuits to be diagnosed. In practical application, the circuits to be diagnosed should be divided into several sub circuits with fewer unknowns in order to avoid the problem of insufficient number of independent equations of fault circuit. In the meanwhile, the circuits should be divided into several independent cut sets and each cut set should be a sub circuit so as to avoid loop circuits in the sub circuits.
4. Component connection model (CCM) method [7, 9, 10, 13]: The idea of CCM is that the element characteristics and its connection in the system are described respectively by two equations: element characteristic equation and its connection equation. The effect concerning the element and the connection to the circuit performances can be discriminated after the element equation and its connection equation are separated.

Table 1 Parameter identification methods

Methods	①	②	③	④	⑤	⑥	⑦	⑧
Combined criteria	Y	YP	N	L and NL	Y	N	Y	N
Iterative	Y	YP	Y	L and NL	Y	Y	Y	
Quadratic programming	Y	YP	Y	L	Y	Y	Y	
L1 norm	Y	YP	N	L and (NL)	Y	Y	Y	
Symbolic transfer function	Y	YP	Y	L	Y	Y	Y	
Voting	Y	YP	N	L	Y		Y	Y
Minimum standard deviation	Y	YP	N	L	Y	Y	Y	
Potential function	Y	YPCS	N	L	Y		Y	
Impact function	Y	YPCS	Y	L	Y	Y	Y	Y
Navid transfer admittance	Y	YP	Y	L and (NL)	Y	Y		
Adjoint circuit	Y	YP	Y	L/NL	Y	Y	Y	
CCM	Y	YP	Y	L/NL	Y	Y	Y	
Equivalent circuit conversion	Y	YP	Y	L	Y			

① Detection; ② Location (Component (P), Circuit (C), System (S)); ③ Identification; ④ Linear (L)/Nonlinear(NL); ⑤ Single fault; ⑥ Multi-fault; ⑦ Soft fault; ⑧ Hard fault

5. Equivalent circuit conversion method [5, 11]: It is a generalization of Y-Δ transformation method, which can transform a star network with controlled source to a polygonal network. The idea of this method is to use the equivalent transform gradually eliminate all the inaccessible nodes in the circuit network, and finally get an network with all node accessible. Then use the test values to determine the parameters of each element of the new network, and use the inverse transform to restore the inaccessible nodes of the original network gradually. Finally find all the element parameters of the original network.

Table 1 shows some comparison of the applicability of these methods.

3.2.2 Network Decomposition Methods

Network decomposition method [5, 7] (or network tearing method, network splitting method): The idea of this method is as follows: Decomposes the large-scale circuit into many sub circuits in some accessible nodes, then uses the self-testing and mutual testing conditions to test these sub circuits, and introduces the logic function to analysis the results of each different test conditions, so as to determine the sub circuit free from faults. Then, divide those sub circuits, which are faulty or not sure, into smaller circuits, and repeat these up steps including verification and logical analysis for the new sub circuits until the smaller fault circuit is diagnosed.

Table 2 Network decomposition methods

Methods	①	②	③	④	⑤	⑥	⑦	⑧
Branch tearing	Y	YC		L/NL	Y	Y	–	–
Node tearing	Y	YP		L/NL	Y	Y		

① Detection; ② Location (Component (P), Circuit (C), System (S)); ③ Identification; ④ Linear (L)/Nonlinear(NL); ⑤ Single fault; ⑥ Multi-fault; ⑦ Soft fault; ⑧ Hard fault

The network decomposition method include branch tearing method, node tearing method, etc.

1. Branch tearing method [5]: For this method, the linear fault diagnostic equations which are based on Kirchhoff's circuit laws (KCL) are used and the voltage testing in nodes under the given current excitation are conducted to divide the circuit into sub circuits or appropriate groups. Then through the consistency test of the fault diagnostic equations, the location of fault circuits can be determined.
2. Node tearing method [38, 39]: For this method, the accessible nodes, and also the inaccessible nodes whose voltage can be calculated according to the network topology and Kirchhoff's current equation are regarded as the tearing nodes. Then the fault sub circuit can be found by the continuity of the Kirchhoff equation in tearing nodes and the positioning technology of fuzzy sets for further location of the fault components.

Table 2 shows some comparison of the applicability of these methods.

3.2.3 Fault Verification Methods

Fault verification methods [5]: This method assumes that there are multiple faults appeared at the same time, and limited measurement and appropriate calculation through a certain algorithm (such as a test of compatibility for some linear equations) can be conducted to determine the fault. This method is based on the fact that the probability of simultaneous faults of all components is minimal to realize multiple fault diagnosis. The way is to inspect whether a subset of the network fails to identify the circuit fault, so the test points decline significantly [10]. Fault verification methods include K-fault diagnostic method, multi-port network independent response vector theorem method, multi frequency transfer function method, maximum rank inverse matrix method, the sensitivity matrix method, transfer impedance matrix method, fault location in linear circuit with tolerance, failure bound method, class fault diagnostic method, etc.

1. K-fault diagnostic method [10]: In this method, the component parameters will be uniformly converted into currents, while the number of faults in the network (not necessarily the fault component number) is limited in k. Therefore, the parameters to be verified are the currents. That the current is zero means the component is fault-free, otherwise the component is considered faulty.

This method has an important assumption: The effect of several failures may not be offset each other. K-fault diagnostic method include, branch diagnostic method, node diagnostic method, and cut-set diagnostic method, etc.

- Branching diagnostic method [9, 10, 14]: For this method, the branches whose fault currents are not zero are directly verified to locate the fault branch and components. If the K branches have a loop, only no more than K-2 fault branches can be determined; and if there is a loop or a cut set in the fault branch set, the method cannot be used; if there is only one branch, who belongs to a loop or a cut set, in the fault branch set, the method has multiple solutions.
- Node diagnostic method [9, 10]: For this method, the nodes that the faulty current is not zero is firstly verified to locate the fault nodes, then the fault branch (element) can be determined by the fault node currents. The advantage of this method is a small amount of post-test calculations and few test points are required.
- Cut sets diagnostic method [2, 14]: For this method, the circuit topology requirement is not restrict, and few accessible nodes are required. This method is flexible, and is applicable to various types of faults. Its diagnostic equations are linear. The fault sources other than the parameters are used in fault description for the non-linear element, and the number of fault source is usually less than that of the parameters. It is also applicable to characteristics curves of nonlinear elements, where analytic function methods are difficult to deal with.

2. Multi-port network independent response vector theorem method [5]: This method is a multiple fault diagnostic method for linear nonreciprocal circuit, and the required internal measurement data are small. It is suitable for a microcomputer for its reduced computation complexity.
3. Multi-frequency transfer function method [5]: This method uses a certain algebraic characters which are invariant concerning the transfer function related to the fault component set, and is with the help of multi-frequency testing to achieve fault diagnosis.
4. Sensitivity matrix method [5, 7, 15] and the transfer impedance matrix method [5]: These two methods both require that the coefficient matrix of the fault diagnostic equations is inverse matrix of full column rank, which limits the effectiveness of these methods. Ss a consequence, some faults cannot be diagnosed, that is, when the coefficient matrix of fault diagnostic equation is not full column rank matrix, faults cannot be diagnosed. In this case, the inverse matrix with maximum rank of the coefficient matrix can still be used for fault diagnosis.
5. Fault location in linear circuit with tolerance method [5]: This method is based on statistical theory. First it assumes that parameters of normal circuit components follow normal distributions to establish the fault circuit model and the

fault diagnostic equation. Then the theorem of necessary conditions for the fault branch set is proposed to achieve multi-fault location. For the single fault diagnosis, the concept of component separation boundary is proposed to narrow down the range of the fault location. Only some measured voltage values of accessible port are needed, and the majority of calculations can be carried out before test, moreover the method can diagnose the soft faults with small deviations.

6. Failure bound method [7, 9–11]: This method assumes that the maximum number of the faults is bounded. The amount of calculation is smaller than that of the K-fault diagnostic method for a large network. The idea of this method is as follows: Let the number of the fault parameters (or the number of fault elements) of a certain network to not exceed a given limit (the limit is concerned with the number of the accessible nodes), and assume that any of the two faults are not mutually offset in the network.
7. Class fault diagnostic method [7, 16, 17]: This method is based on the equivalence relationship of the K element set. All the k element set in the circuit need to be classified and the compatibility of each k elements set (but not all) needs to be inspected so as to diagnose the fault classes. This method is a sub-network-level diagnostic method whose calculation after test is small. And there are no limits on the topology, accessible node locations and incentives.

Table 3 shows some comparison of the applicability of these methods.

Table 3 Fault verification methods

Methods	①	②	③	④	⑤	⑥	⑦	⑧
Branching diagnostic	Y	YP	Y	L/NL	Y	Y	Y	–
Node diagnosis	Y	YP	Y	L/NL	Y	Y	Y	
Cut sets diagnosis	Y	YP	Y	L/NL	Y			
Multi-port network independent response vector theorem diagnosis	Y			L		Y		
Multi-frequency transfer function method	Y	Y		L		Y		
Sensitivity matrix	Y	YP	Y	L	Y	Y	Y	
Transfer impedance matrix	Y	YC		L		Y		
maximum rank inverse matrix	Y	YC		L		Y		
Fault location in linear circuit with tolerance	Y	YP		L	Y	Y	Y	
Failure bound	Y	YP	Y	L/NL	Y	Y	Y	
Class fault diagnosis	Y	YP	Y	L	Y		Y	

Detection; ② Location (Component (P), Circuit (C), System (S)); ③ Identification; ④ Linear(L)/Nonlinear (NL); ⑤ Single fault; ⑥ Multi-fault; ⑦ Soft fault; ⑧ Hard fault

Table 4 Artificial intelligence methods

Methods	①	②	③	④	⑤	⑥	⑦	⑧
Wavelets transform	Y	YP	Y	L		Y	Y	
Multidimensional information fusion	Y	YP	Y	L	Y	Y	Y	Y
ANN	Y	YP	Y	L	Y	Y	Y	Y
SVM	Y	YP		L/NL	Y		Y	Y
Fuzzy	Y	YP	Y		Y		Y	
Genetic algorithms	Y	YP	Y	L	Y		Y	
Fractal theory	Y	YP	Y	L/NL	Y			

① Detection; ② Location(Component (P), Circuit (C), System (S)); ③ Identification; ④ Linear (L)/Nonlinear(NL); ⑤ Single fault; ⑥ Multi-fault; ⑦ Soft fault; ⑧ Hard fault

3.2.4 Artificial Intelligence Methods

Artificial intelligence methods: These methods partially resolve the problems with ambiguity and uncertainty of fault diagnosis and other problems that the conventional methods cannot solve, and they are suitable for fault diagnosis of nonlinear systems. There are many kinds of artificial intelligence methods, such as wavelets transform [9, 18, 28, 40], multi-dimensional information fusion [7, 19], artificial neural network (ANN) [13, 20–22], support vector machine (SVM) [9, 23–27], fuzzy theory [22, 28, 29], genetic algorithms [18, 41], fractal theory [18, 31, 32], etc. These methods are familiar for people in recent years, so here will not illustrate the basic ideas of these methods, just show the comparison of the applicability of these methods as Table 4.

4 Fault Diagnostic Methods Used in BIT

BIT is an abbreviation of Build In Test. According to the definition of MIL—STD-1309-c, BIT is the automatic testing ability of fault-detection and fault-isolation which is provided by the inside of systems and equipments. It means that the system or the equipment itself has an automatic test ability of fault detection, isolation, or diagnosis [33]. Compared to traditional automatic test equipment (ATE), BIT equipment can complete the online detection and location for the system, and effectively reduce the test cost and maintenance time. In the process of BIT design, the selection of fault diagnostic methods and integrating them in the system are need to be considered, and also, the different characteristics and their respective diagnostic method of the analog circuit and digital circuit need to be considered. For digital circuits, the boundary scan technique is often used to diagnose device level and board level system, and formed the relevant criteria and standards [34]. For analog circuits, because the nonlinear circuit has not a general

fault diagnostic method, designers usually determine test points, test parameters and thresholds according to the specific circuits, and combine with fault localization strategy to achieve fault diagnosis and isolation [35]. For most practical circuits—digital-analog hybrid circuits, in some literatures, the testability models are used by abstracting the actual circuits as a logical model for board-level system without using a specific fault diagnostic method. Then a testability analysis is used to obtain fault detection and location strategy; while for digital /analog components, the designers usually depend on the component functions to determine the fault diagnostic methods and criteria. For the system which the digital circuits and analog circuits can be separated, the fault diagnostic methods with digital and analog circuits can be selected respectively [36]. In recent years, with the continuous development of artificial intelligence techniques, the development of these methods of intelligent BIT technique has become a hot research topic, but there are few related cases [37].

5 Summary and Future Work

In this chapter, fault diagnostic methods for analog circuits are classified, and the ideas of these specific methods are described, and then the applicability of each method is analyzed and presented by tables. An investigation concerning current fault diagnostic methods used in BIT is conducted. Based on the review investigation in this chapter, the future work should focus on the case studies of these methods for analog circuits.

References

1. Li F, Woo PY (2002) Fault detection for linear analog IC-the method of short-circuit admittance parameters. IEEE Trans CAS I 49:105–108. doi:10.1109/81.974884
2. Duhamel P, Rault J (1979) Automatic test generation techniques for analog circuits and systems: a review. IEEE Trans CAS 26:411–440. doi:10.1109/TCS.1979.1084676
3. Bandler JW, Salama AE (1985) Fault diagnosis of analog circuits. IEEE Proc 73:1279–1325. doi:10.1109/PROC.1985.13281
4. Sun YC (1989) Analog circuit fault diagnostic theory and Methods. J Dalian Maritime Univ 15:68–75 (in Chinese)
5. Zou R (1989) Principles and methods of analog circuit fault diagnosis. Huazhong University of Science and Technology Press, Huazhong (in Chinese)
6. Yang SK, Guo Y (1990) Review of analog circuit fault diagnostic method. J Hunan Univ (Natural Science) 17:35–40 (in Chinese)
7. Tang RH (1991) Automatic fault diagnosis analog electronic system. Higher Education Press (in Chinese)
8. Liu Y (2008) Methods of analog circuit fault diagnosis—estimation method. J Modern Comm Ind 20:242–243 (in Chinese)
9. Yan S (1993) Fault diagnosis and reliability design of analog system. Tsinghua Univ Press 1:188–196 (in Chinese)

10. Zhou YF (1989). Analog circuit fault diagnosis. National Defense Industry Press, Beijing (in Chinese)
11. Zhao GN, Guo YS (1991) Analog circuit fault diagnosis. Harbin Institute of Technology Press, Harbin (in Chinese)
12. Salama A, Starzyk J, Bandler J et al (1984) A unified decomposition approach for fault location in large analog circuits. IEEE Trans CAS 31:609–622. doi:10.1109/TCS.1984.1085558
13. Robotycki A, Zielonko R (2000) Piecewise linear circuit diagnosis based on component connection model
14. Sun YC (1990) Cut sets method for nonlinear circuit fault diagnosis. J Electron 18:30–34 (in Chinese)
15. Cherubal S, Chatterjee A (1999) Parametric fault diagnosis for analog systems using functional mapping. IEEE DATE, pp 195–200. doi:10.1109/DATE.1999.761121
16. He Y, Tan Y, Sun Y et al (2003) Class-based neural network method for fault location of large-scale analogue circuits. IEEE ISCAS.CAS, pp 733–736. doi:10.1109/ISCAS.2003.1206417
17. Sun YC (1990) Analog circuit fault diagnostic theory and methods. J Commun 5:004 (in Chinese)
18. Ouyang HZ, Liao XB, Liu H et al (2008) Overview of analog circuit fault diagnostic methods. J Electron Sci Technol 21:75–80 (in Chinese)
19. Feng ZH, Lin ZG, Wang W et al (2008) Fault diagnosis of analog circuit based on information fusion. J Anal Electron Meas Instrum 22:47–53 (in Chinese)
20. Yao WJ, You ZH (2004) Fault diagnosis of analog circuits based on neural network. J Central South Univ Nat (Natural Science Edition) 23:50–53 (in Chinese)
21. Yang XF, Fang H (2006) Overview of neural network method for analog circuit fault diagnosis. J Comput Meas Control 14:564–566 (in Chinese)
22. Xin XH, Yang XF, Fu LP (2005) Overview of modern analog circuit fault diagnostic methods. J Autom Instrum 2:1–4 (in Chinese)
23. Grzechca D, Rutkowski J (2009) Fault diagnosis in analog electronic circuits-the SVM approach. Metrol Meas Syst 16:583–598
24. Tang JY, Shi YB (2009) Analog circuit fault diagnosis using fuzzy support vector machine. J Electron Meas Instrum 23:7–12 (in Chinese)
25. Sun YK, Chen G, Li H et al (2008) Analog circuitf ault diagnosis based on testability analysis and support vector machine. J Sci Instrum 29:1182–1186 (in Chinese)
26. Sun YK, Chen G, Li H et al (2008) Application of support vector machine in analog circuit fault diagnosis. Electron J Meas Instrum Sci 22:72–75 (in Chinese)
27. Shou S, Wang Y (2001) Fault diagnosis of nonlinear systems based on support vector machine. J Control Decis 16:617–620 (in Chinese)
28. Yang SY, Hu M, Wang H et al (2008) Research on analog circuit soft fault diagnosis. J Microelectron Comput 25:1–8 (in Chinese)
29. Zhang WP (1991) Research on analog circuit fault diagnostic fuzzy method. J North China Univ Technol 3:67–76 (in Chinese)
30. Milne R, Chandrasekaran B (1986) Fault diagnosis and expert systems. In: The 6th International workshop on expert systems and their applications, pp 603–612
31. Li W, Li TW, Wang GJ, Yi CT et al (2010) Fractal characteristics of analog circuit fault diagnostic method. J Chinese Test 36:14–17 (in Chinese)
32. Li W, Li TW, Wang GJ, Yi CT (2010) Fractal characteristics of analog circuit fault diagnostic method. J Chinese Test 36:14–17 (in Chinese)
33. Shao WL, Zheng WR (2011) Development and application research on BIT technology. J Foreign Electron Meas Technol 30:23–25 (in Chinese)
34. Wen XS, Liu KC (1999) Research status and development trend of board-level BIT based on boundary scan. J Aviat Metrol Meas Technol 19:38–41 (in Chinese)
35. Chen XZ (2008) Design of radar built in test (BIT) system. J Electron Meas Technol 31:134–137 (in Chinese)

36. Chu XJ, Zhang NQ, Zhang X et al (2003) Design of airborne radar BIT. J Avion Technol 34:23–27 (in Chinese)
37. Guai M, Mei XS (2009) Intelligent BIT fault diagnosis system based on BP neural network. J Comput Digital Eng 37:54–56 (in Chinese)
38. Starzyk JA, Liu D (2002) A decomposition method for analog fault location. IEEE ISCAS CAS 3:157–160. doi:10.1109/ISCAS.2002.1010184
39. Chen XJ, W SX, Dai YS (2004) A new analog circuit fault location method. J Northeast Univ Electr Power 24:31–34 (in Chinese)
40. He Y, Tan Y, Sun Y et al (2004) Fault diagnosis of analog circuits based on wavelet packets. In: IEEE TENCON, pp 267–270. doi: 10.1109/TENCON.2004.1414408
41. Zhou L, Yi G (2006) Analog circuit fault diagnosis with tolerance based on genetic algorithms. J Modern Electron Technol 29:121–123 (in Chinese)

An AcciMap Analysis on the China-Yongwen Railway Accident

Lu Chen, Yuan Zhao and Tingdi Zhao

Abstract An AcciMap is a multi-layered causal diagram that arranges the various causes of an accident in terms of their causal remoteness from the accident. This chapter applied an AcciMap to analyse the China-Yongwen railway accident for a more comprehensive view of the accident. Some improvement measures were proposed to prevent similar accidents in the future on the base of this analysis. With the focus of the railway conditions, railway stations, dispatching office, the train drivers and the maintenance personnel, an AcciMap analysis was performed to describe the entire accident trajectory and assemble the contributing factors into a coherent causal diagram that illustrates the interrelationships between them. As a result, some new causes were recognized compared with other analysis methods based on STAMP. Through this study, more critical points and suggestions are provided for enhancing the safety management of Motor Train Unit.

1 Introduction

On 23 July, 2011, a railway accident occurred on the Yongwen High-Speed railway line in Zhejiang province, which took away 40 people's lives. The High-Speed train D301 rear-ended the D3115 with a speed of 99 km/h, as a result, seven cars ran off the rails and two of them went off the bridge [1]. It is considered to be the most serious railway accident in the development of Chinese railway history.

L. Chen
School of Mechanical-Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029, People's Republic of China

Y. Zhao (✉) · T. Zhao
School of Reliability and System Engineering, Beihang University, Beijing 100191, People's Republic of China
e-mail: happy_life03@163.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_105

It is now generally accepted that accidents represent a complex systems-phenomenon, in which causal factors reside at all levels of complex socio-technical systems, and interact across them [4, 2]. The railway accident is a typical complex systems-phenomenon.

Three accident causation models currently dominate the Human Factors literature [3, 4] risik management framework (AcciMap), Reason's [5] omnipresent Swiss Cheese model and Leveson's [2] Systems Theoretic Accident Modelling and Processes model (STAMP). These models have engendered their own distinct approach for analysing accidents. Rasmussen's framework [1] is helpful in this regard, because it has public policy implications for how to design a "vertically integrated" system that can safeguard public health in face of unanticipated events and environmental stressors. For in-depth analysis of single, large scale, complex accidents, the present analysis suggests that the AcciMap method is the most suitable [3] so an AcciMap analysis was performed to describe the entire accident trajectory and assemble the contributing factors into a coherent causal diagram that illustrates the interrelationships between them.

In order to present a more comprehensive and intuitive view of the China-Yongwen railway accident, this chapter applied the AcciMap which is a multi-layered causal diagram that arranges the various causes of an accident in terms of their causal remoteness from the accident.

In this chapter, the factors are separated into five layers corresponding to the domains under which they occur. The results are displayed in a (discrete-mathematics-type) graph, with "nodes" (boxes) representing the factors, and "directed edges" (arrows) representing the causal influence. Furthermore, some critical points and suggestions are provided to enhance the safety management of Motor Train Unit.

2 The Method and Accident

2.1 AcciMap Tool

Graphic representation is always useful in providing an overview of intricate events and processes during accidents. An AcciMap is a multi-layered causal diagram that arranges various causes of an accident in terms of their causal remoteness from the accident [4]. This approach is useful to structurally analysis hazardous work systems, and identifies the interactions in a social-technical system in which accidents unfold themselves [1, 8]. Unlike other methods, this approach assembles the contributing factors into a coherent causal diagram that illustrates the interrelationships between them, thereby highlighting the problem areas that should be addressed to prevent similar accidents in the future.

The socio-technical system actually involved in the control of safety is cross discipline. Policy, decision theory, industrial engineering, human factors, traditional engineering, etc. are all involved in AcciMap. The causal factors of the

socio-technical system are then assembled into a diagram that reveals how conditions and events throughout the system interacted with one another to produce the final negative outcome. This process is useful for highlighting the organizational and systemic inadequacies that contributed to the accident, so that attention is not directed solely towards the events and human errors that led directly to the accident.

2.2 The China-Yongwen Railway Accident

On 23 July, 2011, the accident occurred on the Yongwen High-Speed railway line in Zhejiang province southeast of China. The following is the accident process:

At 19:39, the watch keeper in Wenzhou south station noticed the red light strip, the failure that all lights displayed as red, covered three occlusive sections on the down direction of Yongjia Station. The red light strip indicated that the occlusive section was occupied by a train or in a failure state. Then he reported the problem to the train dispatcher in Shanghai railway administration and also informed the servicemen to do inspection and maintain the failure.

At about 19:45, the servicemen of the signalling branch started to deal with the problem. They maintained some devices without putting the equipment out of service.

At 19:51, D3115 arrived at Yongjia Station.

At 19:54, the train dispatcher commanded the watchman of Wenzhou south station to change the driving mode from Decentralized autonomous control to Unconventional station control. And at 19:55, Yongjia Station changed to Unconventional station control mode, too. Unconventional station control mode means the failure of the interval signal. However, in order to meet the demand of efficiency, the station needs to maintain part of driving by artificial control.

At 20:09, the train dispatcher told the train driver of D3115 that the Yongjia Station is on the Unconventional station control mode, thus, the train should be switched to the on-sight driving mode and continue running rather than stop as a result of signal break down. The driver confirmed this with the train dispatcher.

At 20:12, D301 stopped at Yongjia Station waiting for the signals.

At 20:14, D3115 left Yongjia Station. And at 20:17, the train dispatcher informed the D3115 driver that he should switch to the on-sight driving mode and drive at a speed less than 20 km/h. However, D3115 triggered the Automatic Train protection (ATP) and stopped as the result of the failure of the track circuit. Worse still, from 20:21 to 20:28, the D3115 driver had failed to drive in on-sight mode 3 times, and the communication among D3115 driver, the train dispatcher and the watchman in Wenzhou south station had failed from time to time.

At 20:24, D301 left Yongjia station heading for Wenzhou south station.

At 20:29:26, D3115 succeeded in starting the train by switching to the on-sight driving mode.

At 20:29:32, D301 entered the faulted occlusive section. The driver of D301 saw the slowly moving D3115 and launched emergency brake.

At 20:30:05, D301 with the speed of 99 km/h rear-ended D3115 with the speed of 16 km/h.

The error came from the signal light played a crucial role in the rear-end collision between D301 and D3115. But beyond that, there are also other latent factors existed for a long time which contributed to the likelihood of the initiating event and intermediate events.

3 Accident Analyses with AcciMap

3.1 AcciMap Construction

There's no direct weakness in governmental policy, so that, this chapter didn't consider factor of the ministry of railway in the China-Yongwen railway accident AcciMap. Based on the process of AcciMap and some principle materials, the construction of the accident AcciMap is presented in the Fig. 1 Pertinent preconditions and events are identified and correlated in a graphical way.

3.2 Conditions Summary

The conditions for the accident can be generalized into following texts:

- Thunderstorm weather of the day;
- Maintenance personnel worked without attaching importance to safety standards;
- There existed obstacle in the communication between the train driver and dispatcher;
- Under the abnormal situation, the decision made by the dispatcher might have some error;
- The defect of the management system. The dispatcher and the monitoring personnel in the station had not monitored the abnormal traffic at every moment.

3.3 AcciMap Evaluation

The flow of important boxes is elaborated by the text underneath.

Level 1: dispatching office

The dispatcher of Shanghai railway bureau who related to the accident had neither monitored the abnormal driving section at every moment nor contacted with D3115 train driver. What's worse, the dispatcher neither knew clearly about his responsibility, nor performed the work specification seriously.

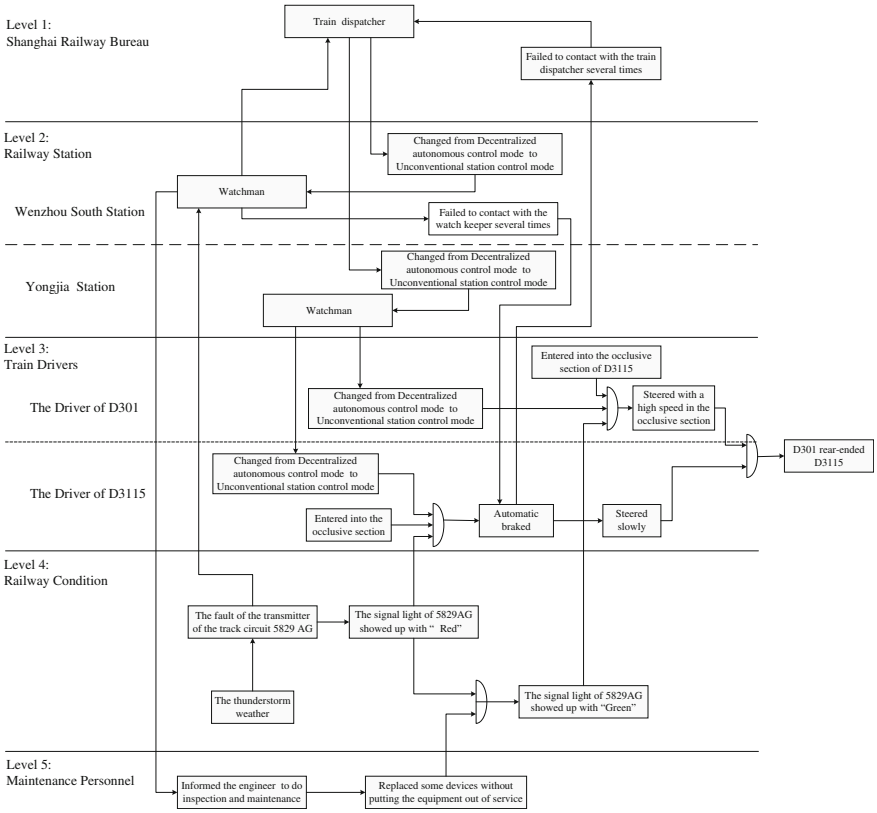


Fig. 1 AcciMap of Yongwen railway accident

The dispatcher was not fully understood about the whole situation of the safety traffic process. In the handling of the emergency, the decision made by the dispatcher might have some error. In the process of the decision-making about train scheduling, if the dispatcher had not approved the D301 train departure until the D3115 train driven out of the 5829 AG section, this accident would have been avoided.

Level 2: railway stations

The monitoring personnel in Wenzhou south station had neither monitored the abnormal driving section at every moment nor contacted with D3115 train driver timely. What's worse, the monitoring personnel neither knew clearly about his responsibility, nor performed the work specification seriously.

Level 3: the train drivers

There existed obstacle in the communication among the D3115 driver, dispatcher and the monitoring personnel in Wenzhou south station. The D3115 driver had called the train dispatcher 6 times and the watchman in Wenzhou south station had called the D3115 driver 3 times but all failed due to communication failure.

It's not the fault of D301 driver. Based on the system design principles and the driving rules, if the signal light showed up with "Green" in the occlusive section, which indicates that there was no train over this section, the train should continue to travel with a high speed.

Level 4: railway conditions

Specific equipments related to the accident were failure. It's comparatively easy to control every specific activity. In the design of the equipment, the technologist should attach importance to the robustness of the device, and improve the reliability of equipment.

Level 5: the maintenance personnel

Electricity workers' safety consciousness was weak, and they didn't explicitly understand the work responsibility. In the maintenance process, electricity workers replaced some devices without putting the equipment out of service. In the repair of signal system, the operation of the electricity workers could lead to the signal light showed up with "green" by mistake, which leads to the driver of D301 who rely deeply on the signal system continued to travel with a high speed mistakenly.

3.4 Compared with Other Analysis

Adopting AcciMap to analyse the accident process, some new causes were recognized compared with other analysis methods based on STAMP [7].

- Over the level 1 of the AcciMap construction, i.e. dispatching office, the decisions made by the dispatcher had some logic errors. After the emergency appears, the dispatcher should know well about each bug might occur in the process of train running, for the sake of safety. The train travels from Yongjia station to Wenzhou south station only need 10 min normally. From the perspective of "safety first", if the dispatcher had not approved the D301 train departure until the D3115 train driven out of the 5829 AG section, this accident would have been avoided.
- The actors at each level were not committed to safety over the whole process of the train's safety running. For instance, the maintenance personnel did not know well about the train operation process, so that they had replaced some devices without putting the equipment out of service. Actors at each level were unable to see how their decisions interact with those made by actors at other levels, so the threats to safety were not obvious before an accident occurs. If actors at each level had a clear understanding of the whole process of the train's safety running, the accident would be avoided to a large extent.

4 Discussions and Suggestion

As the authors' intent of AcciMap is improvement of the system, it can be drawn from the figure that the possible links and factors are sensitive to improvement. Through the analysis of the China-Yongwen Railway Accident by AcciMap, the following experience is to be kept so as to prevent similar accidents from occurring in the future.

1. The information interaction was not available. For example, during the "Red light strip" section, D3115 fail to call the train dispatcher several times, otherwise, the accident can be avoided.
2. After the failure occurs, the design on the interaction logic in the failure treatment process of the system emerged some flaws. For instance, after the "Red light strip" appeared, if the dispatcher had not approved the D301 train departure until the D3115 train arrived at Wenzhou south railway station, this accident would have been avoided.
3. Ensure that the actors at each level have the required competencies.
4. Ensure that the actors at each level are committed to safety (i.e. each level have a positive safety culture).
5. Make the work objectives at each layer explicit.

In order to prevent similar accidents, it is necessary to find the root causes at the different levels of socio-technical system. The AcciMap is one of useful methods for identifying the root causes of accident. The AcciMap could be added to train travelling safety management system structure to support the system to be improved.

References

1. Dennis MW, Kim JV (2003) Sociotechnical systems, risk management, and public health: comparing the North Battleford and Walkerton outbreaks. Reliab Eng Syst Saf 80:253–269
2. Leveson NG (2004) A new accident model for engineering safer systems. Saf Sci 42 (4):237–270
3. Paul MS et al. (2012) Systems-based accident analysis methods: a comparison of AcciMap, HFACS, and STAMP. Saf Sci 50:1158–1170
4. Rasmussen J (1997) Risk management in a dynamic society: a modelling problem. Saf Sci 27:183–213
5. Reason J (1990) Human error. Cambridge University Press, Cambridge
6. State Administration of Work Safety (2011) The investigation report on the "7.23" Yongwen line major railway accident. (in Chinese). http://www.chinasafety.gov.cn/newpage/Contents/Channel_5498/2011/1228/160577/content_160577.htm. Accessed 2011
7. Song T, Zhong DM (2012) A STAMP analysis on the China-Yongwen railway accident. The 31st International Conference, AFECOMP 2012, Magdeburg, Germany
8. Sun LL (2008) Adapting the AcciMap for an analysis of the SQ006 accident. The 26th congress of international council of the aeronautical sciences, Anchorage, Alaska

Studying the Potentials of Physical Asset Management of Hybrid Base Stations in Telecommunication Companies

Nikola Asurdzic and Macro Macchi

Abstract The importance of energy related topics has been increased during last years, and it will be in future for sure. In particular energy cost is estimated to increase; further on one of the hot topics today is also the global climate change, which is due to increased CO_2 and other greenhouse gases emissions. To reduce these impacts decrease of usage of fossil fuels can be a proper lever, and can be obtained both by improving energy efficiency and by using renewable energy resources. On the other hand, focusing specifically in the Telecom sector by introduction of Android, iPhone, iPad, Kindle and social network such as Facebook, demand for cellular data traffic has grown significantly. Indeed, Base Transceiver Stations (BTS) consume a maximum portion of the total energy used in a cellular system (around 60 %). Eventually, it is known that Information and Communication Technology (ICT) already represents about 2–2.5 % of total carbon emissions and this is expected to increase every year. The most convenient ways to reduce energy consumption of BTS is usage of renewable energy sources (wind and sun).This is recognized by existing theory and practice. Available literature covers the performances of Hybrid Base Station (HBTS), site indicators, on one side, and, on the other side, the necessity of the Telecom Company to reduce energy consumption and GHG emission on sustainable way. These are the two extremes of this knowledge domain: more precisely, on one side, there is a literature related to the vendors of HBTS and, on the other side, there is literature dedicated to Telecom Companies as technology users. But there is no bond and, in fact, insufficient knowledge technology management. Indeed what we see as a missing link here is the availability of a clear information and design of the practical "guideline" for implementation of HBTS in a Telecom company. The literature of Physical Asset Management (PAM) offers good references for an empirical research to unveil such missing link. In particular, based on the holistic lifecycle approach, that is a relevant concept of PAM. We planned to develop an on field research with Telecom

N. Asurdzic (✉) · M. Macchi
Politecnico di Milano, Milan, Italy
e-mail: nikola.asurdzic@polimi.it

M. Macchi
e-mail: marco.macchi@polimi.it

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_106

companies, to understand how the technology is managed in practice. On-going case study provided some insights to us during writing this paper. Hence, the research will aim at assessing every single step of HBTS implementation in a Telecom Company. This approach will cover all the Asset Life Cycle stages, ranging from Concept and Project Approval to Decommissioning and Disposal, so to let emerge existing practices in a broad range of activities. This paper will focus on the whole architecture established for the empirical research, focusing initially on the literature survey used as background, which may help then showing the structure of the questionnaire used for the research and its first results after an initial validation with Telecom experts.

Keywords Telecommunication · Physical asset management · Renewable energy sources

1 Introduction

The importance of energy related topics has been increasing during the last years. Energy cost is estimated to grow in the future. Furthermore, it is not merely an economic matter: energy consumption and global warming potentials are amongst the most relevant environmental impact categories, and enhanced attention on such issues is demonstrated by the ever strict regulations from governments and public awareness on "green" performance. Henceforth, indicators for environmental impacts are nowadays a subject of more attention: when dealing with a technical/technological solution, it is a matter of fact that economic factors are not the only drivers, also the environmental factors are becoming essential; therefore, the energy impacts of technical/technological solutions are being studied in terms of energy requirements (measured e.g. in kWh), subsequent effects on greenhouse gas (GHG) emissions, etc.… It is indeed relevant that an industrial or service activity is assessed for such impacts, with the purpose to drive the contribution of industrial/service sectors to sustainable development.

As public utilities, telecommunication companies (shortly, in the remainder, telecom) should and, indeed, pay attention on such relevant matters. On the other hand, telecom companies are clearly operating in a growing market [1] in particular, due to ICT (Information and Communication Technology) evolution and, in particular, the introduction of Android, iPhone, iPad, Kindle and social networks such as Facebook and Twitter, the demand for cellular data traffic has grown significantly. The market growth is obviously representing a positive issue as economic factor, even if it can be a source of negative effects for the environmental burdens. This paper discusses the trade-offs between the economic and environmental factors: to this end, the results of a literature review on current trends in energy and GHG emissions are summarized (Sect. 2), both assuming a global and a specific (telecom oriented) perspective; a market analysis is also provided (in Sect. 2)

considering the unprecedented growth of new demand for voice and data transfer in the telecom sector; the effects of such relevant trends lead to focus on the needs for an enhanced technology management in the telecom companies (Sect. 3), having a special insight on the adoption of Physical Asset Management (PAM) as relevant lever in order to manage the introduction of new technologies. The particular concern kept within the paper aims at especially looking at the introduction of "green" technologies: to this end, Hybrid Base Station (HBTS) are analysed as a specific solution for telecom using renewable energy sources (i.e., HBTS are powered by wind and sun).

The paper is driven by a relevant question on the background: "*are the management processes of a telecom company prepared for enabling effective the implementation of "green" technologies, such as it is the case of HBTS?*"; "*how should management processes be changed in order to enable a more affordable and effective implementation of "green" technologies (that is HBTS)?*"; more precisely, "*how is decision making enabled, considering different needs in terms of supporting resources—information, knowledge/expertise, procedures …—along the asset (HTBS) life cycle?*". All in all, it is interesting to focus on the acquisition, in-service support and disposal of the assets (HBTS) in a telecom company, considering that supporting resources are dispersed in the telecom business context: the evidences of market analysis in fact reveal that such supporting resources are fragmented amongst many operators present in the market; in other terms, a telecom company—managing the assets with the basic objective to achieve the quality of the provided service—is in a sense just "the end of the line", whilst other actors are involved, especially OEMs (Original Equipment Manufacturers) of physical assets, as the HBTS. As such, there is a relevant information and knowledge to be mobilized and captured by telecom companies, sourced from third parties: to this end, PAM seems a relevant lever that can be considered as an interesting subject of further study.

2 Background and Motivations

2.1 Global Energy Consumption

The importance of energy related topics has been increasing during last years, and it will be in future for sure. According to IEA (International Energy Agency) data from 1990 to 2008, the average energy use per person increased 10 % while world population increased 27 %. Regional energy use also grew from 1990 to 2008 (see Table 1).

Energy consumption in G20 increased by more than 5 % in 2010 after the slight decline of 2009. In 2009, world energy consumption decreased for the first time in 30 years, by −1.1 % (equivalent to 130 mega tonnes of oil), as a result of the financial and economic crisis, which reduced world GDP by 0.6 % in 2009 [2]. China became the world's largest energy consumer (18 % of the total) since its consumption surged by 8 % during 2009 (up from 4 % in 2008). Oil remained the largest energy source (33 %), despite the fact that its share has been decreasing over

Table 1 Regional energy use (kWh/capita & TWh) and growth 1990–2008 (%)

	kWh/capita			Population (million)			Energy use (1,000 TWh)		
	1990	2008	Growth (%)	1990	2008	Growth (%)	1990	2008	Growth (%)
USA	89,021	87,216	−2	250	305	22	22.3	26.6	20
EU-27	40,240	40,821	1	473	499	5	19.0	20.4	7
Middle East	19,422	34,744	79	132	199	51	2.6	6.9	170
China	8,839	18,608	111	1,141	1,333	17	10.1	24.8	146
Latin America	11,281	14,421	28	355	462	30	4.0	6.7	66
Africa	7,094	7,792	10	634	984	55	4.5	7.7	70
India	4,419	6,280	42	850	1,140	34	3.8	7.2	91
The world	19,422	21,283	10	5,265	6,688	27	102.3	142.3	39

Source IEA/OECD, Population OECD/World Bank

time. Coal posted a growing role in the world's energy consumption: in 2009, it accounted for 27 % of the total consumption.

It is well known that resources of fossil energy are limited, and they are running out of capacity since humans are excavating each day an enormous amount of them. Regulations and technologies are leading to slow down the usage of fossil fuels, but they cannot stop excavation. There is a need for a sustainable, long term solution. The implementation of renewable ways for collecting the energy is a solution fostered for the future: wind, solar, biofuels, geothermal, waves are just some of the sources providing energy in a sustainable way [3].

2.2 Global CO_2 and GHG Emission

Since 1751 approximately 337 billion tons of carbon have been released to the atmosphere from the consumption of fossil fuels and cement production. Half of these emissions have occurred since the mid'70ies. The 2007 global fossil-fuel carbon emission estimate, 8365 million metric tons of carbon, represents an all-time high, and a 1.7 % increase from 2006 [4].

As more people consume more fossil fuels, the increasing amounts of carbon dioxide (CO_2) emitted into the atmosphere have begun to dramatically change atmospheric composition. At the beginning of the industrial revolution, CO_2 concentrations in the atmosphere were 280 parts per million (ppm). Today, the concentrations have reached 390 ppm (see Table 2). With increased CO_2 concentrations implicated as a major factor in global warming and ozone layer, understanding the relationship between emissions and economics has become a crucial aspect worldwide. The United Nations Framework Convention on Climate Change in fact calls for "*stabilization of greenhouse-gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system*" [5].

One of the most convenient way of reducing CO_2 and GHG emissions, but still to fulfil demand for energy, is to switch from fossil fuels energy resources to renewable energy resources such as wind, sun and waves power, geothermal power, hydropower and biofuels [3].

Table 2 Current greenhouse gas concentrations

Gas	Pre-1750 tropospheric concentration	Recent tropospheric concentration	Absolute increase since 1750	Percentage increase since 1750 (%)
Carbon dioxide(CO_2)	280 ppm	392.6 ppm	112.6 ppm	40.2
Methane (CH_4)	700 ppb	1,874 ppb	1,174 ppb	167.7
Nitrous oxide (N_2O)	270 ppb	324 ppb	54 ppb	20.0
Tropospheric ozone (O_3)	25 ppb	34 ppb	9 ppb	36

2.3 New Demand for Voice and Data Transfer in Telecom Sector

Privatization swept the world and high competition was introduced in many countries, especially in mobile telecom market. The movement from state monopolies to privatized telecom companies, the introduction of competition where none existed before, and the rise of mobile service motivate a comprehensive examination of the changing worldwide telecom business.

During the last decade, there has been a particular, tremendous growth in cellular networks. The number of subscribers and the demand for cellular traffic has escalated astronomically. By August 2007 there were 3 billion mobile connections in the world. Considering that it took fixed line telephony more than 125 years in order to reach 1 billion lines, this is clearly a dramatic growth. Indeed, mobile connections increased from 1 to 3 billion in less than 6 years. At present, new mobile phone connections are growing at around 45 million/month or, equivalent to say, 1.5 million/day [6].

With the introduction of Android and iPhone devices, the use of e-book readers such as iPad and Kindle and tablet PCs, and the success of social networking giants such as Facebook and Twitter, the demand for cellular data traffic have also grown significantly in the recent years. Henceforth, mobile operators have been developing while meeting these new demands in wireless cellular networks, competing on new services and reduced costs [1] (Table 3).

Future demand for data transfer in telecom market can be even increasing due to the introduction of Internet of Things (IoT). As an idea, IoT is a novel paradigm that is rapidly gaining ground in the scenario of modern wireless telecoms. The basic idea of this concept is the pervasive presence around us of a variety of things or objects—such as Radio-Frequency Identification (RFID) tags, sensors, actuators, mobile phones, etc.—which, through unique addressing schemes, are capable to interact with each other and cooperate with their neighbours in order to reach common goals. Since most of the things that will be "a piece of pie" of IoT are

Table 3 Mobile cellular subscriptions (per 100 people)

Country name	1992	1997	2002	2007	2011
Austria	2	15	83	119	155
Italy	1	21	94	151	158
Switzerland	5	15	79	109	131
USA	4	20	49	82	93
Brazil	–	3	19	64	124
Mexico	–	2	25	61	82
Japan	1	31	64	85	105
UAE	2	12	75	143	149

Source The World Bank

mobile it is necessary to use telecom networks to connect and interact with them. Wireless, mobile connections represent the most suitable solution for networking in order to support functioning of IoT [7].

2.4 Energy Consumption in Telecom Sector

Unprecedented growth in cellular industry has pushed the limits of energy consumption in wireless networks. The price that has to be paid for this enormous growth in data rates and market penetration is a rising power requirement for ICT systems. Both in server farms as core units of the internet [8] as well as in mobile communications systems [9], a rise of the power consumption of 16–20 % per year can be observed in the last years, corresponding to a doubling every 4–5 years.

There are currently more than 4 million base stations (BSs) serving mobile users, each consuming an average of 25 MWh/year. The number of BSs in developing regions is expected to almost double by 2012 [1].

Other interesting numbers are the followings: today, more than two thirds of the operational expenses in cellular networks in India are due to the buying of diesel for running the generators of cellular BSs; cellular network operators spend beyond 200 million euro each year for their electricity bills in Germany alone; electricity has grown to a cost factor comparable to the total wages of the engineers who keep the network running. As a further example, Fig. 1 outlines the increase of energy consumption in one of the leading Telecom Company in Europe, Vodafone: blue pillars represent energy consumption of Vodafone D2, a branch in charge for German territory, and red pillars show energy consumption in the whole Vodafone group in a period of 3 years.

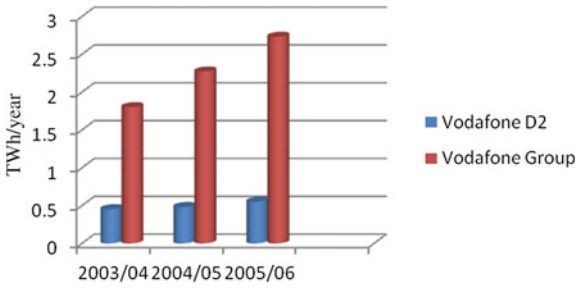


Fig. 1 Electricity consumption of Vodafone network. (*Source* Vodafone Corporate Social Responsibility Reports 2001/02 to 2005/06)

2.5 CO_2 and GHG Emission from Telecom Sector

ICT already represents about 2–2.5 % of total carbon emissions, presently accounting for approximately 0.86 metric gigatonnes of carbon emissions annually. In particular, fixed-line telecoms account for about 15 % of the total ICT carbon emissions, while mobile telecoms contribute an additional 9 %, LAN and office telecoms about 7 %. Servers, including cooling systems, account for 23 %. From [10] lifecycle assessment studies, and other published data sources, it is estimated that approximately 0.14 % of global CO_2 emissions and approximately 0.12 % of primary energy use are attributable to mobile telecoms. This compares with 20 % of CO_2 emissions and approximately 23 % of primary energy use for travel and transport, for example. Using other "tangible" numbers, it is worth underlining that: (i) the annual CO_2 footprint of the average mobile subscriber is around 25 kg—which is comparable to driving an average car on the motorway for 1 h, or running a 5 W lamp for a year; (ii) many telecom operators have about the same energy consumption today as they did in 1995, but with twice as many total subscribers.

Therefore, technology improvements have been relevant to keep energy usage very low, even if there are still great opportunities for the ICT industry to reduce CO_2 emissions. For example, smart use of telecom, intelligent homes and offices, and travel substitution could all have a dramatic effect on energy usage and CO_2 footprint. Nonetheless, if nothing is done, the ICT contribution to GHG emissions is projected to nearly double—to about 4 %—by 2020 [11]. ICT usage is in fact expected to expand rapidly over the coming decade, especially in developing countries [11].

3 Network, Critical Assets and Existing Solutions

3.1 The Mobile Network and Its Critical Assets for Energy Consumption

The typical mobile network consists of three main components: (i) Core Network (for switching, interfacing to the fixed network, billing, etc.); (ii) Base Station (to support radio frequency interface between network and mobile terminals); Mobile Terminals (used by mobile subscribers to make phone and data calls). Figure 2 provides a summarizing overview of such an infrastructure.

In terms of energy consumption, the key components are the base stations, as the number of base stations is relative high with relative high energy consumption per station. The energy consumption of a typical single base station could vary from 0.5 kW up to 2.0 kW. As number of core networks is low, even one single core network could consume energy up to 10 kW. On the other hand, the energy consumption of mobile terminals is very low due to mobile nature. Figure 3 shows a

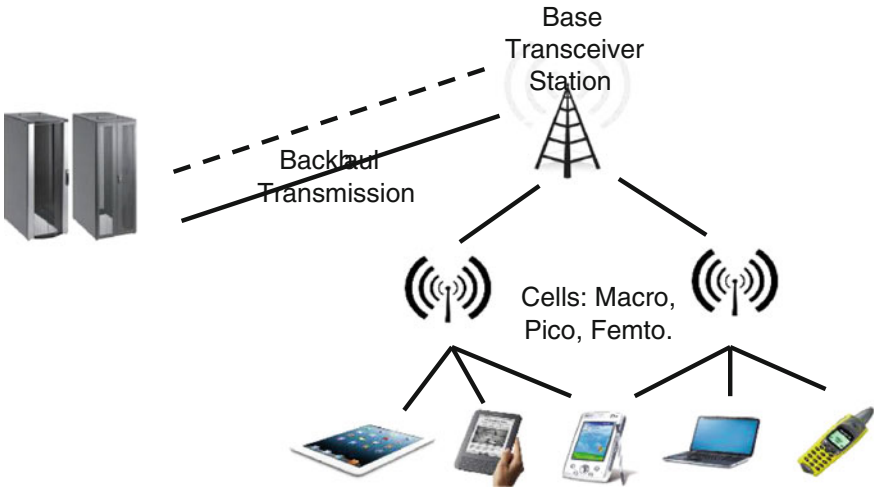


Fig. 2 Mobile network and its components

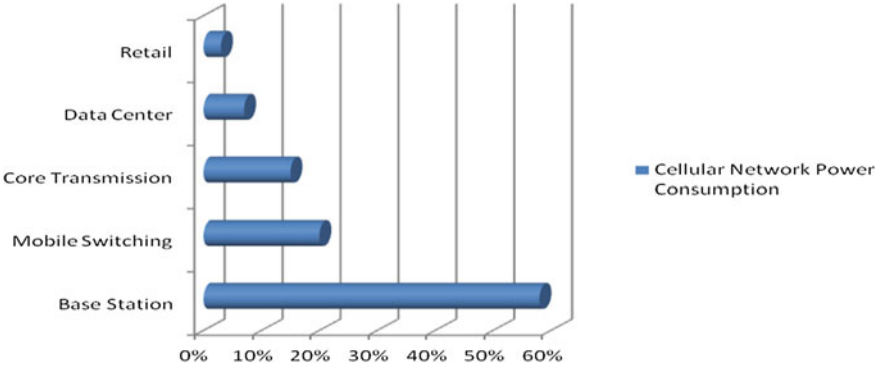


Fig. 3 Breakdown of power consumption in a typical mobile network: Power consumption of a typical wireless mobile network [17]

breakdown of power consumption in a typical cellular network, giving an insight into the possible critical assets for reducing energy consumption in wireless communications [12].

3.2 Solutions to Reduce the Energy Consumption of Base Stations

An important goal is then to reduce the energy consumption of base stations, with the purpose to reduce total energy consumption of mobile networks thus having a sustainable net effect by combining decreased costs and carbon footprint.

The solutions to decrease energy consumption of base stations, and thus to reduce cost and CO_2 emissions, could be divided into three main categories: (i) minimizing the based stations energy consumption; (ii) minimizing the number of base stations sites; (iii) usage of renewable energy sources by means of Hybrid Based Stations (HTBS) powered by wind and sun. Minimizing BTS energy consumption is more a technical on-site solution, which includes both software and hardware improvements. Minimizing the number of BTS sites is based on an optimized design solution precedent to the creation and implementation of architecture of telecoms' network. The last way of reducing energy consumption and CO_2 emissions in Telecom sector, and the focus of this paper, is the usage of renewable energy sources through HTBS.

3.3 The Use of Renewable Energy Sources as a Solution

Renewable energy is energy derived from resources that are regenerative. For this reason, renewable energy sources are fundamentally different from fossil fuels, and do not produce greenhouse gases like CO_2. The most feasible renewable energy source for BTS sites are considered solar and wind. In most cases also a hybrid solution combining solar and wind resources is actually the most feasible for autonomous BTS sites [10, 13–15].

For BTS sites the renewable energy could be used for several reasons like: in case of long distance to the electricity grid, with an unreliable grid, and with the purpose to reduce amount of CO_2 emission and energy consumption [1].

Typical Hybrid Base Transceiver Station (HBTS) consists of: wind generator situated on the top of the mast which is high up to 10 m, photovoltaic cells covering up to 20 m^2 surfaces and battery pack which is used during the night or during no-wind hours. This battery pack is charged by wind and solar power when conditions are good. Alternatively, HBTS can include: power diesel generator or hydrogen fuel cells. Both this energy sources are used alternatively when the weather conditions are not satisfactory for producing sufficient amount of energy from wind and sun and when battery is in low power (see Telecom Power Solutions, Power Oasis, http://www.power-oasis.com/).

According to a new report from [16], annual deployments of off-grid power supplies, using renewable or alternative energy sources, for remote mobile base stations will grow from fewer than 13,000 worldwide in 2012 to more than 84,000 in 2020. All in all, more than 390,000 of such base stations will be deployed from 2012 through 2020, the study concludes. Currently, there are around 640,000 off-grid base stations which represent 16 % of all number of worldwide base stations (total number of 4 millions).

4 Studying the Potentials of Physical Asset Management

4.1 Motivation and Structure of the Questionnaire for Empirical Research

Existing literature and practice in the field of HBTS can be split on two different sides: on one side, there is the manufacturer/vendor of HBTS and HBTS auxiliary equipment, on the other side there is the Telecommunication Company. Indeed, manufacturers/vendors literature is mostly concerned on a number of items such as: performances of HBTS, indicators of specific site, performances of HBTS that are customized for the specific site. On the other hand, Telecom literature is much wider in scope, because it discusses the necessity of energy consumption and GHG emission reduction, as well as the necessity to cover off-grid areas. Manufacturers/vendors and Telecommunication companies are the two extremes of a knowledge/information domain, and there is no clear bond between the two sides, which may eventually result in insufficient knowledge/information for technology management. Indeed, what we see as a missing link is the availability of clear information, together with the design of a practical "guideline" for implementation of HBTS in the context of a Telecom company. The present paper will focus on the architecture established for the empirical research with the purpose to study such a knowledge/information domain; the PAM literature is considered as a background for its justification.

More specifically, the architecture is based on the main hypothesis that the PAM approach should cover such a knowledge/information domain. To this concern, seven stages of development of HBTS in a Telecom company are assumed; further on, an on-going case study is carried on in order to help proving that these stages are recognized in a Telecom Company during implementation of HBTS. Table 4 is a derivation from some PAM literature references, mainly from [17], and it distinguishes the seven stages; Fig. 4 shows a breakdown of Life Cycle Cost in CAPEX and OPEX, relating costs to the different stages of the life cycle, as derived from [17, 18].

Having in mind that we have to fulfil all stages mentioned above in order to support HBTS implementation, we formed the questionnaire. Structure of the questionnaire consists of two parts: the first part has the aim to collect general information on respondents such as company name, country of origin, function and sector of corresponding person; the second part is reserved to questions more related to the adoption of PAM approach. On the whole, there are 58 questions divided in seven PAM stages and three levers: strategy, management and technology. Such levers can be motivated considering the general literature of PAM and, in particular, different statements revealing how PAM is intended. We can take out for example the statement from [17], which is well representing: "*Asset management is concerned with applying technical and financial strategy and sound management practices to deciding what assets we need to meet our business aims, and then to acquiring and logistically sustaining the assets over their whole life, through to disposal*". Table 5 provides a summary of the questionnaire, with the number of questions through each PAM stage and lever.

Table 4 PAM stages in the life cycle of a single HBTS

Name of the stage	Description of the stage
Feasibility stage	The initial engineering investigations and preliminary design work to establish project costs associated with operational requirements, environmental requirements, safety case requirements, policy requirements and submission of total capital requirements for financial approval
Project approval stage	Review of the financial justification, and analysis of business risk prior approval, or not, of capital expenditure
Detailed design stage	The development of all engineering calculations, engineering specifications, electrical/mechanical/civil/structural drawings, equipment selection specifications, maintenance strategy, maintenance documentation, construction specifications, and contractual documentation needed to build, commission and operate the HBTS and its equipment
Procurement stage	The detailed specification and purchase of equipment and materials to construct and start-up the operation of HBTS
Construction (installation, commissioning)	This consists of site works, earthworks, services, materials and labour to build a single HBTS, divided in installation and commissioning
	Installation—the placing and mounting of equipment into position on their foundations and the provision of services and process interconnections to operate them at design specification to provide quality of the service at the required company availability and service provision rate
	Commissioning—testing and proving the HBTS equipment can be operated safely and would meet all design specifications prior to acceptance and handover
Operations	This consists of safely and sustainably running HBTS equipment to provide in-specification service at the design rate and cost
Decommissioning	This refers to the safe removal from service of company and equipment for sale, storage or scrapping

4.2 Using the Questionnaire for State of the Art and State of Practice Analysis

The questionnaire may be used to confirm what is publicly available at the state of the art, as well as to reveal what is effectively done at the state of practice in a company. To this regard, it is worth observing that a subset of questions can be answered also based on public literature: hence, a respondent of the questionnaire may "customize" the answers, reflecting the specific attitude of his/her company as well as comparing the answers with respect to what generally known as the state of the art. The remaining answers have to be collected necessarily from the Telecom Companies and Original Equipment Manufacturers because of their "practical nature", not precisely known from literature review: that is, they are related to the

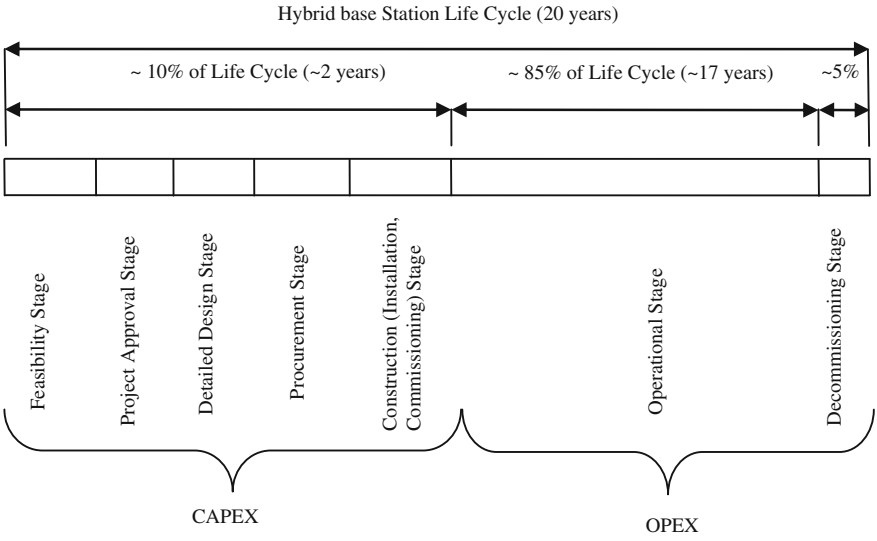


Fig. 4 Breakdown of Life Cycle Cost for HBTS in a given timeline (20 years of expected life)

state of practice. Subsequently, it can be said that answers are hidden in knowledge and information domains of different business actors operating in the telecom market. Table 6 outlines the number (percentage) of questions that can be answered also using existing literature as an initial source, and the number (percentage) of questions that requires answers from specific respondents, either from the Telecom Company or from the Original Equipment Manufacturer. This classification has been done by leaning on PAM methodology from the reviewed literature. Regarding this, we can see that some stages and sub-stages in PAM consist of the universal activities for the most of the fixed assets and, in this case, can be also applied to HBTS asset management. While, on the other hand, some of the stages are totally dependent of the experience of focused company and has to be observed directly during on-going case study.

From Table 6 we can see that number of questions that can find answers in literature is slightly minor than the number of questions that seek for the answers in the state of practice: 28 versus 30. This reveals our expectation that PAM approach of HBTS may come out from an adequate balance between state of the art and state of practice. These numbers in fact motivate the questionnaire and, in particular, the collection of answers on the practical side, to achieve a comprehensive picture of PAM of HBTS. Further on, it is worth observing that the number of questions from the same lever is quite similar (comparing literature and practical side), which implies that none of the levers can be obtained from literature only.

Structure of the questions in our questionnaire is mainly based on PAM knowledge/information. That implies that these questions are not yet perfectly suited and shaped for Telecom Company, they are more general to all fixed assets.

Table 5 Number of questions in specific PAM stages and levers

	Physical asset management stages						
	Feasibility stage	Project approval stage	Detail design stage	Procurement stage	Construction (installation, commissioning)	Operations	Decommissioning
Strategy	8	2	4	3	1	1	2
Management	2	0	7	1	0	16	2
Technology	1	1	0	3	3	1	0
	Total number of questions						58

Table 6 Cataloguing the questions according to the source

	Also literature	Only respondent
Strategy	10 (17 %)	11 (19 %)
Asset management	13 (22 %)	15 (26 %)
Technology	5 (9 %)	4 (7 %)
Total	28 (48 %)	30 (52 %)

On-going research in Telecom Company will enable customizing and shaping these questions and bringing it closer, more focused on the specific issues due to the Telecom Sector, while still following the PAM approach.

4.3 Initial Results from the State of Practice Analysis in a Case Study

The first feedbacks of application of the questionnaire in a case study confirmed our assumption, based on the existing literature, that the implementation of HBTS in a Telecom Company is conducted using a similar methodology as PAM methodology through the stages we mentioned earlier in this paper (see Table 4). Further on, some answers collected during the case study provided us with new insights: these revealed that a Telecom Company may conduct some procedures in different stages of PAM, differently from what we assumed initially in our questionnaire; these feedbacks will be used to enhance the questionnaire for the next application in other Telecom Companies.

Results from the case study also showed that the strategy of implementation of HBTS may be carried out in parallel technology implementation phases. In the specific company target of study, the first phase concerned only the implementation of standard base station equipment not related to the power section (i.e., antenna, tower, cables, shelter, cooling system, switches and monitoring system), while the second phase consisted of the implementation of Hybrid part of the Base Station equipment (photovoltaic cells, wind turbine, special batteries for downtime support, controller for optimization usage of energy resources and special monitoring system for hybrid part of the equipment, in this case called "Greenmeter" www.greenmeter.com). First phase was conducted on the regular organizational regime, as for any other standard base station, since equipment and technology is unchanged comparing to non-hybrid base station: since this is one of the primary Telecom Company activities, the company's technical personnel is well trained and experienced for standard implementation. The second phase, i.e. the introduction of Hybrid part of the equipment, was new to the Telecom technicians: a special project team was then formed and charged for HBTS implementation; this team consisted of the employees from different departments such as design, maintenance, implementation, technical and site design area. Main source of knowledge for this phase for the team was existing literature and consultation with OEM. After completion of project

documents, tender was opened for the procurement of equipment and selection of manufacturers. OEM for each part of the Hybrid equipment was then selected based on the optimal ratio of technical characteristics and prices. Beyond procurement, we also found out, during our analysis, that the criticality for operations and maintenance of the hybrid part of the equipment was judged as medium by the Company; besides, we got insight in the list of standards that the company must follow during the construction stage of the HBTS (construction, electro technical, and lightning rod).

We also got the answers to some questions that have a "customized nature", in the sense that they should reflect the unique attitude of each company in its business context. As an example, the answer to the question; "describe how your company decides to approve, or not, the financing of Hybrid Base Station acquisition, by identifying the relevant drivers (financial versus non-financial)". The answer of this question is unique for each company, because it depends on its strategy in the competitive context. In the case study, policies and interests of the company were the initial cornerstone for project initiation. Since this Company operates in a country, where the price of the electricity is one of the lowest in Europe, reduction of OPEX of HBTS was not so feasible. Hence, what initiated the HBTS implementation project was the marketing department, which identified the HBTS as a great opportunity for promoting the "Green" attitude of the company. This is unique example, and we can of course consider that the reason for financial approval for HBTS implementation will vary from company to company, from country to country, from region to region and so forth.

5 Concluding Remarks

As concluding remarks, it is worth pointing out that the questionnaire proposed in this paper is a result of our design choices after literature analysis. Further on, it can be used as indicator of our perception of the partial results one can have only from literature, poor on some particular aspects of the asset life cycle management. Accordingly to such perception, more than 50 % of answers should represent practical experience of each individual company, therefore they would be collected directly at company level to effectively comprehend how the stages of PAM, i.e. from feasibility until decommissioning and disposal, are actually run.

With all the answers from the questionnaire we would expect to achieve an insight on HBTS implementation guided by usage of PAM, integrating in a balanced way state of the art from literature and state of practice from the point of view of companies operating in the telecom market. Besides, we believe that, with this research approach, we would enhance knowledge and information publicly available for some of the PAM stages and, in that sense, we would highlight issues needed to enhance the whole PAM approach of implementation of HBTS in an "average" Telecom Company. Last but not least, what we see as a final result is a practical "guideline" that will facilitate the implementation of HBTS to Telecom Companies, having the holistic life cycle perspective of PAM.

The literature of Physical Asset Management (PAM) offers good references to drive the empirical research. In particular, based on its holistic lifecycle view, the PAM approach can support Telecom Company endeavour for HBTS implementation. This is the main assumption of the present work. Deriving from PAM well known concepts, we can then identify the target issues of HBTS implementation process that can be enhanced by introduction of our potential design guideline on: (i) the Life Cycle Cost, providing insight through a breakdown into CAPEX and OPEX of HTBS; (ii) the strategy and technology recommended for HTBS use; (iii) the resources and expertise needed to support acquisition and in-service of HBTS; (iv) the systems and facilities needed to support HBTSs throughout their life. The case study analysis, whose initial insights have been provided in this paper, has provided positive feedback on the expected results of PAM as a relevant lever to integrate knowledge/information domain dispersed in the telecom market. Moreover, during this case study we can state that the first implementation of the hybrid part of base station for Telecom Company is rather "unfamiliar" process. Details of each stage of implementation were not all known by technical staff, so they had to consult existing literature and OEM in order to learn and to assume each successive stage of implementation. This fact identifies relevant gaps interesting for our study. Indeed, we believe that "product" of our study, "guideline", will facilitate the implementation of the above-mentioned process by giving insights inside the PAM stages. The project team would not have to spend time reviewing the existing literature and consultation with OEMs; then, a "guideline" will contain all these information in a concise and comprehensive level.

6 Future Opportunities

As was predicted by [19], telecom operators in the developed world will deploy alternative energy solutions at cell sites in three stages:

- First, operators will look to the few areas so remote no grid is available - this already started happening.
- Second, telecom operators will start deploying alternative energy systems in areas where the power grid is available but renewable energy is abundant, such as the sun-drenched regions of the southwest or wind-swept central plains. These will represent the "best-effort" deployments which will use alternative power for the majority of the site's power needs, but with fall back on the power grid when power supplies get low.
- Thirdly and finally, Telecom operators and HBTS companies will get into the power-generation business. As green energy technology gets cheaper and more efficient, sites will be able to store up excess energy, saving some of it for "rainy days", but distributing much of it back into the energy grid. Wireless operators globally are in the unique position of owning highly distributed networks with hundreds of thousands of cell sites, which can act as gigantic distributed energy

farms. All that is necessary to make such a scenario feasible is the right set of economic conditions and incentives.

We see these predictions as future guideline for necessity of implementation of HBTS. All literature points to the fact that in the near future most of resources of energy will be shifted from fossil ones to renewable ones. This is the reason why we see bright future for the "practical guideline" for HBTS implementation that we want to form. In future, usage of renewable energy sources will be necessity, not the technical innovation or cutting edge technology. "There are some favourable conditions and there are some conditions working against operators. Once the return on investment is there, though, you will see this happen" [19]. When this will happen, this would be beneficial also to new PAM applications: accordingly with our first empirical feedbacks, there may be expectations for enhanced PAM methodologies applied to the telecom business.

References

1. Hasan Z, Boostanimehr H, Bhargava VK (2011) Green cellular networks: a survey, some research issues and challenges. IEEE Commun Surv Tutorials 13(4):524–540
2. Enerdata Publications Yearbook 2012 (2012) Global Energy Market Review in 2011
3. Dincer I (2000) Renewable energy and sustainable development: a crucial review. Renew Sustain Energy Rev 4(2):157–175
4. Boden TA, Marland G, Andres RJ (2010) Global, regional, and national fossil-fuel CO2 emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U. S. Department of Energy, Oak Ridge
5. United Nations (1992) United Nations Framework Convention on Climate Change
6. Telenor (2007) Economic Contribution of Mobile Communications. Report
7. Atzori L, Iera A, Morabito G (2010) The internet of things: a survey. DIEE University of Cagliari/University "Mediterranea" of Reggio Calabria/University of Catania, Italy
8. Koomey JG (2007) Estimating total power consumption by servers in the U.S. and the world
9. Vodafone Corporate Social Responsibility Reports 2001/02 to 2005/06
10. Ericsson (2007) Sustainable energy use in mobile communication. White Paper
11. International Telecommunications Union [ITU] http://www.itu.int
12. Louhi JT (2007) Energy efficiency of modern cellular base stations. Telecommunications Energy Conference, 2007. INTELEC 2007, 29th International. IEEE
13. Lubritto C, Petraglia A, Vetromile C, Caterina F, D'Onofrio A, Logorelli M, Marsico G, Curcuruto S Telecommunication power systems: energy saving, renewable sources and environmental monitoring. Department of Environmental Science, II University of Naples/ National Agency for Environment (APAT), Italy
14. Telecom Power Solutions, Power Oasis (2013) http://www.power-oasis.com/
15. Kusakana K, Vermaak HJ Hybrid renewable power systems for mobile telephony base stations in developing countries. Department of Electrical Engineering and Computer Systems, Central University of Technology, Bloemfontein
16. Pike Research (2013) http://www.pikeresearch.com
17. Hastings NAJ (2010) Physical asset management. Springer, London
18. Jötten, Gerrit et al. (2011) ¨Assessment of flexible demand response business cases in the smart grid.' Proceedings of the 21st International Conference on Electricity Distribution CIRED, Frankfurt.
19. Kevin Fitchard (2009) Alcatel-Lucent: First "on-grid" alternative energy cell sites on the horizon

Application of Feature Extraction Based on Fractal Theory in Fault Diagnosis of Bearing

Wentao Li, Xiaoyang Li and Tongmin Jiang

Abstract Fractal theory can be applied to state recognition and fault diagnosis of bearing for the nonlinear property of rotation machinery's vibration signal. In this paper, a feature extraction method based on fractal theory is introduced and the fractal feature is extracted by computing the correlation dimension of vibration signals in different conditions. Correlation dimension can be determined by G-P algorithm and relevant parameters' selection methods are discussed. C–C method is used to calculate the time delay of phase space reconstruction. The example of bearing shows that the correlation of bearing in fault condition is much higher than that in normal condition, which can help to recognize bearing's state and discover bearing's fault promptly.

Keywords Feature extraction · Fractal theory · G-P algorithm · Correlation dimension

1 Introduction

The fault diagnostic technology is applied to recognize mechanical equipment's operation state using information which is generated in the working procedure. However, part of mechanical equipment's vibration signals contain intense nonlinear

W. Li · T. Jiang
Product Environmental Engineering Research Center, Beihang University, 37 Xueyuan Road, Haidian District, Beijing 100191, China
e-mail: liwentaohigh@163.com

T. Jiang
e-mail: jtm@buaa.edu.cn

X. Li (✉)
Science and Technology on Reliability and Environmental Engineering Laboratory, Beihang University, 37 Xueyuan Road, Haidian District, Beijing 100191, China
e-mail: leexy@buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_107

feature in non-stationary conditions. So using traditional time domain or frequency domain approaches can't get accurate result. In contrast, the fractal theory can be applied to vibration signal process.

As an important mathematical branch of nonlinear subjects, Fractal theory has developed rapidly since 1970s. Now there are mainly two ways to apply fractal theory to fault diagnosis of mechanical equipment. The first way is to extract the fractal feature of abrasive dust and provide evidence for on-line fault diagnosis by getting the wear rate indirectly. The second way is to analyse the characteristic signals of machine operation state and calculate the fractal dimension, then the state of machine can be inferred according to fractal dimension. In general, the latter method is mainly used to fault diagnosis of rotating machinery.

2 Fractal Feature Extraction

2.1 Correlation Dimension

Fractal dimension has many definitions and corresponding calculation methods have been developed. Since the collected data signal is often a single time series in condition monitoring and fault diagnosis, correlation dimension is a common way to data analysis. Some research has been done on calculating correlation dimension, such as measuring relational method, correlation function method, spectrum method and G-P method [1]. G-P method is the most widely used algorithm now and it is proposed by [1] in 1983, according to Whitehead imbedding theorem and Packing phase space reconstruction. The basic thinking is as follows:

For a time series x_1, x_2, x_3, …, x_{N-1}, x_N to phase space reconstruction, assume that the embedding dimension is m, we can get the reconstruction space [3],

$$\begin{aligned} X &= [X_1, X_2, \ldots, X_{N0-M}, X_{N0-M+1}]^T \\ &= \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(m-1)\tau} \\ x_2 & x_{2+\tau} & \cdots & x_{2+(m-1)\tau} \\ \ldots\ldots & & & \\ x_{N_m-1} & x_{N_m-1+\tau} & \cdots & x_{N-1} \\ x_{N_m} & x_{N_m+\tau} & \cdots & x_N \end{bmatrix} \end{aligned} \tag{1}$$

where N_m is the number of reconstruction space's vectors, and $N_m = N-(m-1)\tau$, m is the embedding dimension; N is the data number of time series, τ is the time delay. So the definition of correlation integral function is

$$C(r) = \frac{2}{N_m(N_m-1)} \sum_{1 \le i < j \le N_m} H(r - \left\| x_i - x_j \right\|) \tag{2}$$

where $H\{*\}$ is Heaviside function(step function), that is

$$H(X) = \begin{cases} 1 & X \geq 0 \\ 0 & X < 0 \end{cases} \tag{3}$$

So the correlation dimension can be calculated as

$$D_2 = \lim_{r \to 0} [\ln C_m(r) / \ln r] \tag{4}$$

The specific steps to calculate correlation dimension using G-P algorithm is:

(1) For a time series x_1, x_2, x_3, …, x_{N-1}, firstly choose a smaller value m_0 to reconstruct the phase space according to Eq. (1).
(2) Calculate the correlation integral function according to Eq. (2).
(3) For a suitable range of r, get the fitting curve of $lnC(r) \sim ln(r)$ according to Eq. (4).
(4) Increase embedding dimension and let $m_1 > m_0$, repeat steps (2) and (3) till, D_2, the estimated value of corresponding correlation dimension, becomes stable with m increasing. Thus we get the correlation dimension, and the corresponding m is the optimal embedding dimension.

2.2 Relevant Parameters of Phase Space Reconstruction

Since the reconstructed phase space can't fully recover mechanical system's dynamic behavior when we choose the value of m and τ unreasoningly, the key point of phase space reconstruction is how to obtain the value of m and τ. Now there are two ways to get m and τ. one way believes that m and τ are independent, and correlation function approach [4], mutual information approach [5] are used to calculate τ; while FNN [6] approach and Cao approach are used to calculate m. The other way believes that m and τ are related, such as C–C [7] approach. Research now shows that C–C approach is more reasonable to determine m and τ.

When to handle the bearing's vibration signals, this paper uses C–C approach to determine τ, and take the estimation of m as the reference value. Then increase the value of m slightly and we can get the optimal embedding dimension when correlation dimension curve converges (Fig. 1).

Correlation integral function expresses the probability of the distance between arbitrary two points less than r in phase space. We define a test statistic as [3]

$$S_1(m, N, r, t) = C(m, N, r, t) - C^m(1, N, r, t) \tag{5}$$

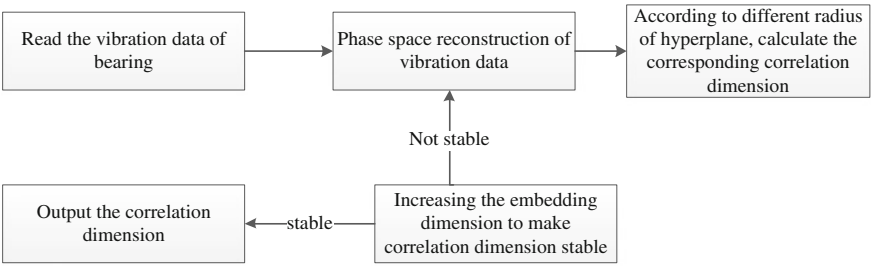


Fig. 1 Flow chart of G-P algorithm

where m, N, r, t are introduced in Sect. 2.1. To compute Eq. (5), we take the blocking strategy and define

$$S_2(m,N,r,t) = \frac{1}{t}\sum_{s=1}^{t}\left[C_s(m,N/t,r,t) - C_s^m(1,N/t,r,t)\right] \tag{6}$$

Let $N \longrightarrow \infty$, we have

$$S_2(m,r,t) = \frac{1}{t}\sum_{s=1}^{t}\left[C_s(m,r,t) - C_s^m(1,r,t)\right] \tag{7}$$

In order to measure the max maximum deviation of $S_2(m, r, t) \sim t$ for all the radius r, we define

$$\Delta S_2(m,t) = \max\left\{S_2(m,r_j,t)\right\} - \min\left\{S_2(m,r_j,t)\right\} \tag{8}$$

Then the optimal time delay, τ_d, is the first zero point of $S_2(m, r, t) \sim t$ or the first local minimum point of $\Delta S_2(m, t)$.

According to BDS statistical conclusions we can get the reasonable estimation of N, m, r. Here we set N = 3,000, m = 2, 3, 4, 5, $r_i = i*0.5\sigma$, $\sigma = \mathrm{std}(x)$ (σ is the standard deviation of time series). i = 1, 2, 3, 4. Calculate Eqs. (9) and (10)

$$\overline{S_2}(t) = \frac{1}{16}\sum_{m=2}^{5}\sum_{i=1}^{4} S_2(m,r_i,t) \tag{9}$$

$$\Delta\overline{S_2}(t) = \frac{1}{4}\Delta S_2(m,t) \tag{10}$$

And the first local minimum point of $\Delta\overline{S_2}(t)$ is τ_d.

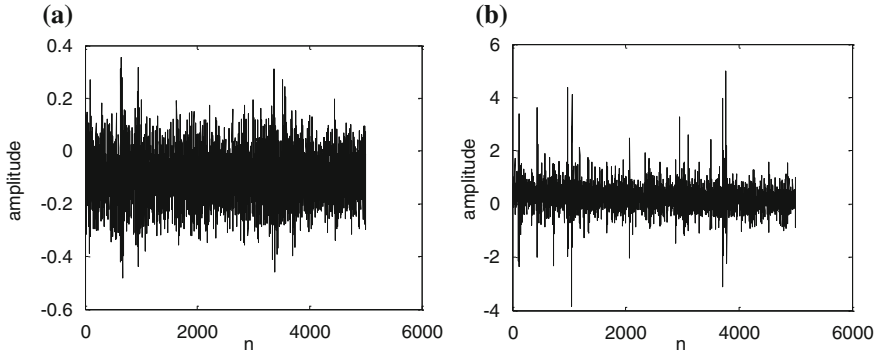


Fig. 2 The vibration signal of condition (**a**) and condition (**b**)

3 Example

In this section, fractal feature extraction of bearing's vibration signal is implemented and the vibration data comes from a bearing life testing carried out by [8]. The Bearing test rig hosts the bearing on a shaft which is driven by an AC motor. The rotation speed was kept constant at 2,000 rpm. A radial load of 6,000 lbs. is added to the shaft and bearing by a spring mechanism. Vibration data was collected every 20 min and data collection is conducted by a National Instruments LabVIEW program. The data sampling rate is 20 kHz and the data length is 20,480 points. At the end of the test a crack was found near the shoulder of a bearing.

Here, two sets of vibration data of the wear bearing are selected to feature extraction. We denote the normal condition of bearing by condition A and the wear condition of bearing by condition B. Considering that computing too many points is unnecessary and time-consuming, we intercept 5,000 points from the 20,480 points for analysis. The corresponding vibration signals are shown in Fig. 2.

From Fig. 2 we can see that the signal amplitude of condition B is much higher than that of condition A. It shows the rapid wear of bearing at the end of test.

Using C–C method to get τ, as is shown in Fig. 3. The $LnC(r)$-$ln(r)$ curve is shown in Fig. 4.

Figure 4a, b show the $LnC(r)$-$ln(r)$ curve with different m in condition A and Fig. 4c, d show the D_2-m curve in condition A and B, which reflects how the correlation dimension changes versus m. When the D_2-m curve converges we get the value of correlation dimension and the corresponding m is the optimal embedding dimension.

The result of vibration signal analysis is given in Table 1. It shows that the correlation dimension in condition B is much higher than that in condition A, which can help us to recognize bearing's state and discover bearing's fault promptly.

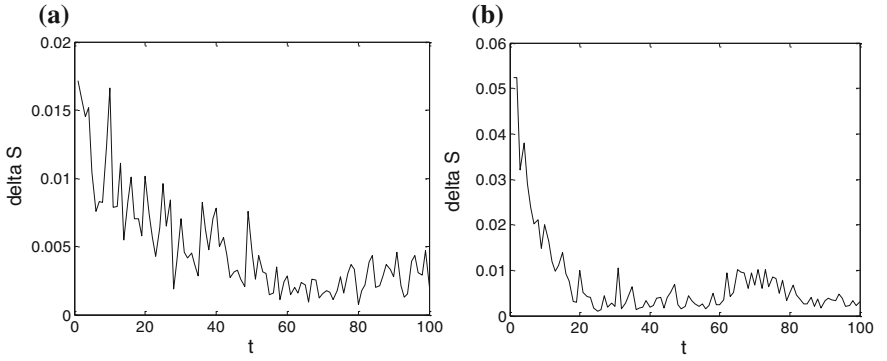


Fig. 3 Calculation of τ using C–C method. $\Delta\overline{S_2}(t) \sim$t curve using C-C method in condition (**a**), $\Delta\overline{S_2}(t) \sim$t curve using C-C method in condition (**b**)

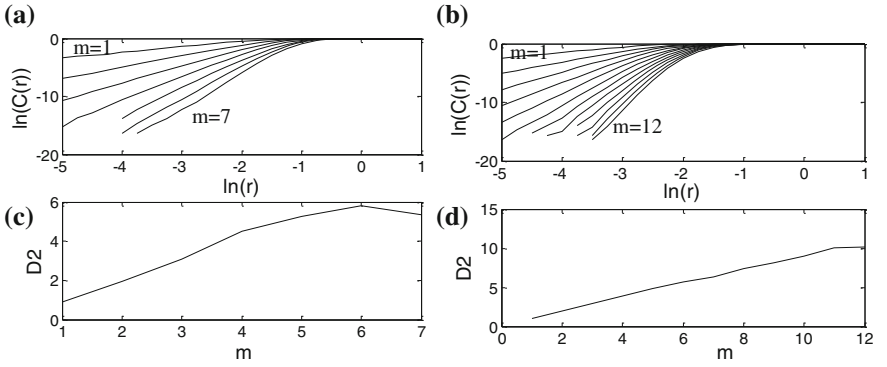


Fig. 4 LnC(r)-ln(r) curve and D2 (correlation dimension)-m curve of condition A and B

Table 1 The bearing's correlation dimension in different conditions

Condition of bearing	τ	m	D_2 (Correlation dimension)
A	3	6	5.5
B	1	11	10.1

4 Conclusion

In this paper a feature extraction method based on fractal theory of is proposed to bearing's state recognition and fault diagnosis. G-P algorithm for calculating correlation dimension is introduced and the selection of related parameters is also

discussed. The example of bearing's fractal feature extraction shows that vibration the bearing has different correlation dimensions in different conditions, and the correlation dimension of wear bearing is much higher than that in normal condition. By analysing the vibration signal, we can recognize bearing's state and discover bearing's fault promptly.

References

1. Qing-hua W, Xing-biao Z (2004) Application of fractal theory to fault diagnosis for hydraulic pump. J Dalian Marit Univ 30(2):40–43
2. Grassberger P, Procaccia I (1983) Characterization of strange attractors. Phys Rev Let t A 50 (5):346–349
3. Lv Jinhu (2007) Analysis and application of chaotic time series. Wuhan University Press, Wuhan, pp 57–66
4. Kantz H, Schreiber T (1997) Nonlinear time series analysis. Cambridge University Press, London, p 127
5. Fraser AM, Swinney HL (1986) Independent coordinates for strange attractors from time series. Phys Rev A (S1094-1622) 33:1134–1140
6. Kennel MB et al (1992) Geometry from a time series and phase space reconstruction using a geometrical construction. Phys Rev Let t A 45(1):25–28
7. Kim HS, Eykholt R, Salas JD (1999) Nonlinear dynamics delay times, and embedding windows. Phys D(S0167-2798) 127:48–60
8. Qiu Hai, Lee Jay (2006) Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. J Sound Vib 289:1066–1090

Performance Monitoring with Application of Reliability Growth Analysis

Allen S.B. Tam

Abstract Reliability growth analysis is a popular tool in monitoring reliability changes over time. The methods used include Duane plots and the Army Materiel Systems Analysis Activity (AMSAA) model. The reliability growth analysis is traditionally applied to the time domain of failure. The amplitude of failure is not considered. For equipment or processes with same MTBF should not be treated with the same priority without looking into loss of opportunity in wealth creation. This research proposes the application of reliability growth method to failure losses in unit of production loss for monitoring changes in magnitude of production loss due to failure events. In conjunction with reliability growth in time domain, a new analytical visualisation of reliability assessment, named Reliability Quadrant Graph, is proposed. This research provides a new way for high level reliability performance monitoring.

Keywords Reliability growth · Reliability quadrant · Equipment reliability analysis · Production losses

1 Introduction

Reliability growth analysis is a popular tool in monitoring reliability changes over time. Popular modules include Duane plot and the AMSAA model. The model has been applied to a number of different industry such as defence [1, 2], power transmission and distribution [3], software engineering and information technology system [4, 5].

(The concepts of this paper was developed during the author's employment at Silcar).

A.S.B. Tam (✉)
SP-AusNet, Level 31, 2 Southbank Boulevard, Southbank, VIC 3006, Australia
e-mail: sbatam@yahoo.com.hk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_108

More recent development and extension in the theory of reliability growth model was reported in the Reliability and Maintainability Symposium (RAMS) in recent years. Tananko et al. [2] applied the growth model to product verification test for mobile gun system during production for improving the design for reliability. Crow [6] proposed the extended continuous evaluation reliability growth model that is to be applied for reliability growth evaluation over a single or multiple test phases for "operational-like" testing. Kraisch [7] pointed out that common practise of additions of test times at failure n reliability growth test data analysis with multiple test items is unsuitable for design or manufacturing issues. Kriasch [7] proposed an analytical method applying the original Non-Homogeneous Poisson Process (NHPP) to correct the error. Strunz and Herrmann [8] addressed the issue with limited data points (such as liquid rocket engine) by introducing a new Bayesian estimation based methodology.

These developments however, focus in the application of the growth model in time domain, monitoring failure time. The amplitude of impact of the failure to business is not covered.

This research proposes a new paradigm of application of the growth model by applying the model to failure losses in unit of production losses. Production losses can be measured in tonnes (mining and manufacturing industry), Mega Watt Hour (power generation), downtime (manufacturing), customer minutes off supplies/services (telecommunication and power transmission and distributions). This allows monitoring of the trend of losses magnitude. In conjunction with reliability growth in time domain, a new analytical visualisation of reliability assessment, named Asset Health Condition Graph, is proposed.

2 Reliability Growth Models

2.1 Time Growth Model

The time based reliability growth model is a non-homogenous Poisson process (NHPP) [9]. This growth model is also known as the AMSAA reliability growth model. The AMSAA reliability growth model in general is given as [10]:

$$N(t) = \lambda t^{\beta} \tag{1}$$

If the model applies to the data, the natural logarithms of equation give:

$$LnN(t) = Ln\lambda + \beta Lnt \tag{2}$$

And by plotting $LnN(t)$ on y-axis versus $Ln(t)$ on x-axis, where:

- t is the cumulative time
- $N(t)$—Cumulative number of failures/events at cumulative time t

- λ—y intercept when t = 1
- B—indicator of reliability improvement

The model is generally used in monitoring system reliability trends using the β value as the health indicator. When β value is:

- More than 1, the system is deteriorating (i.e. decreasing Mean Time Between Failure (MTBF))
- Equals to 1, the system is stable (i.e. constant MTBF)
- Less than 1, the system is improving (i.e. increasing MTBF)

2.2 Loss Growth Model

The reliability growth model is applied to monitoring the time domain of failure but the model by itself is insufficient as the failure consequence in losses amplitude. Frequent small failure that has little impact to production will rank higher than less frequent failure with increasing losses amplitude.

To address this deficiency, this research proposes the application of the growth model to losses magnitude.

Applying the same model but replacing "t" with "l", and to avoid confusion of the slope values, the slope of the model is termed α. The growth model for losses amplitude is therefore:

$$N(l) = \lambda l^{\alpha} \tag{3}$$

$$LnN(l) = Ln\lambda + \alpha Lnl \tag{4}$$

By plotting $LnN(l)$ on y-axis versus $Ln(l)$ on x-axis, where:

- l is the cumulative losses magnitude
- N(l)—Cumulative number of events at cumulative losses magnitude l
- λ—y intercept when t = 1
- α—slope

When α value is:

- More than 1, the mean loss magnitude between events is decreasing
- Equals to 1, the mean loss magnitude between events is constant
- Less than 1, the mean loss magnitude between events is increasing.

The slope α has a different meaning to β. The system is 'good' when the mean losses between events is decreasing, which means that the failure impact to production has a decreasing trend.

3 Reliability Quadrant Graph

The quadrant graph (Fig. 1) is a visualisation tool for assessing plant reliability. On the X-axis is the time growth slope and on the Y-axis is the losses magnitude growth slope. If slope for both loss and time equals to 1, this indicates that there is no change in both time and loss magnitude for the given dataset. The quadrant graph takes the coordinate (1, 1) as the origin, which divided the graph into 4 quadrants. The system/subsystem growth slopes values will fall into one of the 4 quadrants. Each quadrant is an indicator of individual plant system/subsystem reliability and is interpreted as follows:

1. Quadrant (A) $\alpha > 1, \beta < 1$
   - Decreasing mean loss between events
   - Increasing MTBF
   - Good Quadrant
2. Quadrant (B) $\alpha > 1, \beta > 1$
   - Decreasing mean loss between events
   - Decreasing MTBF
3. Quadrant (C) $\alpha < 1, \beta < 1$
   - Increasing mean loss between events
   - Increasing MTBF
   - Failures are becoming less frequent but more severe

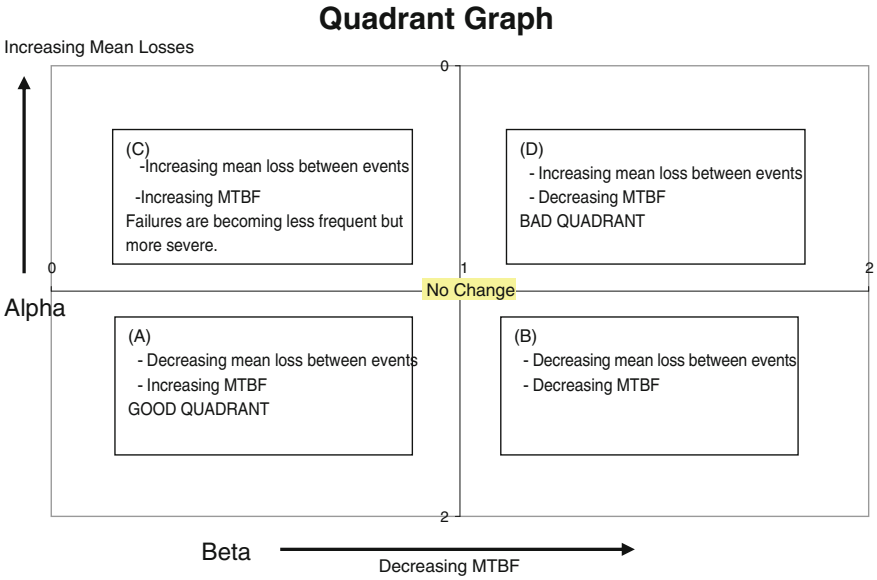


Fig. 1 Quadrant graph

4. Quadrant (D) $\alpha < 1, \beta > 1$
   - Increasing mean loss b/w events
   - Decreasing MTBF
   - Bad Quadrant

If all the plant systems growth slopes values are obtained and plotted on the same graph, this scatter plot (Fig. 2) provides a performance indicator of plant reliability. If most of the points are in:

1. Quadrant (A)
   - Plant reliability is in control
2. Quadrant (B)
   - Systemic issues
   - Quick fixes
   - Breakdown work ok, but has not fixed the root cause
   - Failures becoming more frequent but less severe
3. Quadrant (C)
   - Failures becoming less frequent but more severe
   - One-off major loss events—needs more attention to hidden plant problem

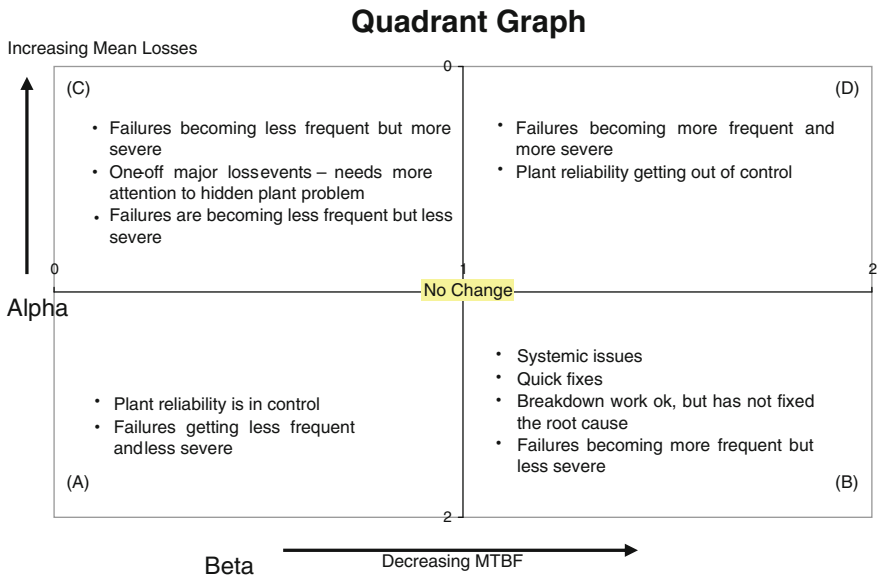


Fig. 2 Quadrant graph—plant wide analysis

4. Quadrant (D)
   - Failures becoming more frequent and more severe
   - Plant reliability getting out of control.

4 Application

4.1 Numerical Example

A system of a manufacturing plant is considered. At failure, system is assumed to undergo minimum repair. Under the current production practices and maintenance policy, a total of 10 failures that causes losses of production are experienced to-date. The time of failure and the total losses of production due to failure are given in Table 1.

The growth slope can be calculated by using the maximum likelihood estimator formula:

$$\hat{\beta} = \frac{n}{n \ln(T^*) - \sum_{i=1}^{n} \ln T_i} \tag{5}$$

- T^*—is the time of last failure for failure terminate analysis, or for time terminated analysis is the future time of interest
- n—Cumulative number of failure
- i—the "i" failure

Table 1 Failure data for a system of a manufacturing plant

Event	Weeks	Losses in unit of production	Cumulative losses	LN (No. of events)	LN (Weeks)	LN (cumulative losses)
1	8	200	200	0	2.0794	5.2983
2	20	100	300	0.6931	2.9957	5.7037
3	30	300	600	1.0986	3.4011	6.396
4	50	250	850	1.3862	3.9120	6.7452
5	60	60	910	1.6094	4.0943	6.8134
6	90	500	1410	1.7917	4.499	7.2513
7	106	20	1430	1.9459	4.6634	7.2654
8	150	200	1630	2.0794	5.0106	7.3963
9	180	100	1730	2.1972	5.1929	7.4558
10	260	300	2030	2.3025	5.5606	7.6157

For this example, the calculated values are:
Time growth:

$$\hat{\beta} = \frac{10}{10\ln(260) - 41.41} = 0.704$$

Losses magnitude growth:

$$\hat{\alpha} = \frac{10}{10\ln(2030) - 67.94} = 1.22$$

5 Case Study

In this section, a power station is considered to illustrate application of the reliability quadrant graph. The numbering used represents a subsystem in the power station.

The two plots, Figs. 3 and 4, compare for the same systems, using two different time period. Figure 3 analyse data between the year 2002–2008. Figure 4 looks at data trend between 2002 and 2009. The additional dataset of 2009 provide comparison of changes of performance within the year. The target of the scatter plot is to have points (representing a system) to move from right to left, and from top to bottom towards the lower left hand quadrant. It is interpreted that as plant systems are moving for an increasing trend of loss magnitude and failure events occurrence to a decreasing trend of loss magnitude and failure events occurrence.

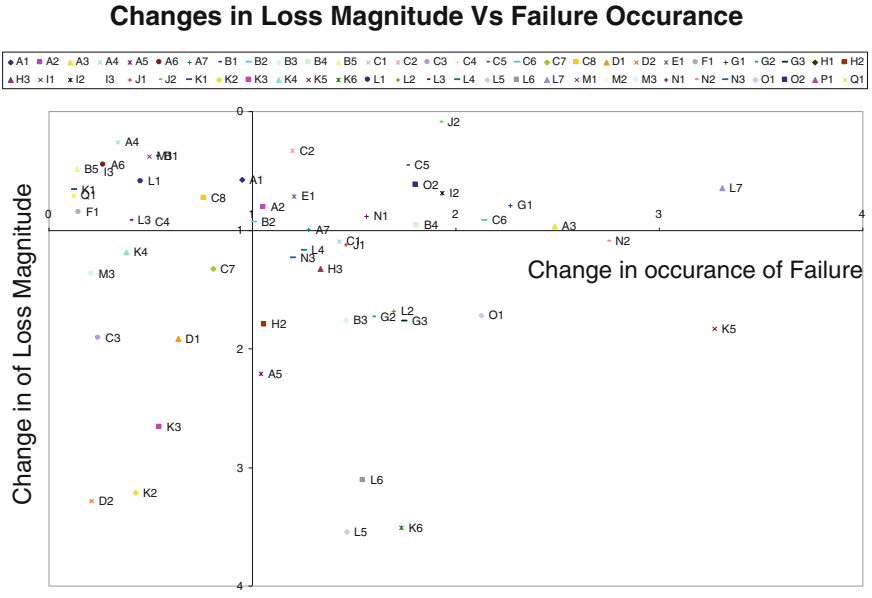


Fig. 3 Reliability quadrant graph (Time Period 2002–2008)

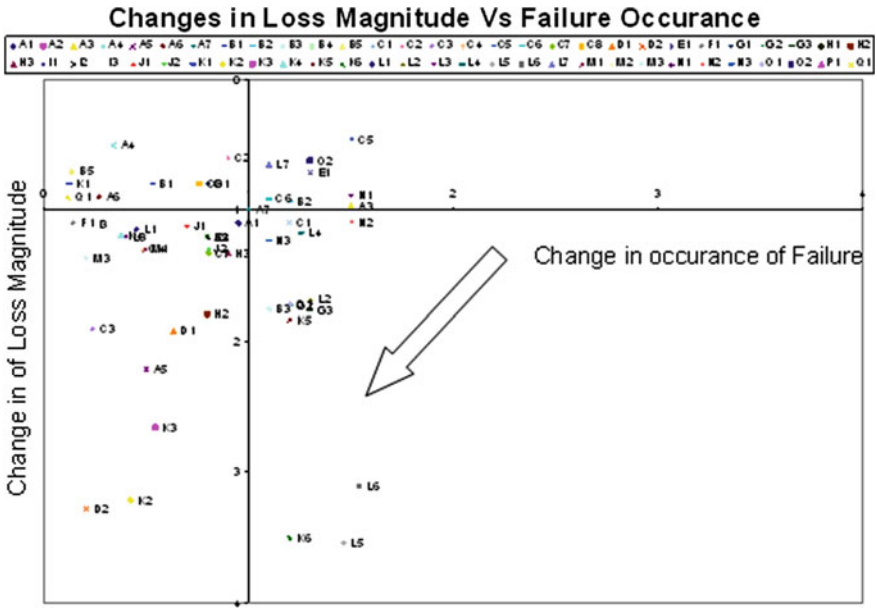


Fig. 4 Reliability quadrant graph (Time Period 2002–2009)

6 Conclusion

This paper proposed a new application of the reliability growth analysis. The proposed reliability quadrant graph provides a high level visualisation for managers and engineers to understand the changes of their asset based on data. The application of this approach provides high level summary of plant reliability performance.

Acknowledgment The author would like to acknowledge the support of Silcar for the development of this paper between 2007 and 2010.

References

1. Hall BJ, Mosleh A (2008) An analytical framework for reliability growth of one-shot systems. Reliab Eng Syst Saf 93:1751–1760
2. Tananko DE, Kumar S, Paulson J, Chang NJ (2009) Reliability growth of mobile gun system during production verification test. In: Proceedings of the reliability and maintainability symposium (RAMS), 2009
3. Shi Q (2010) Research and application of a reliability growth model for 220 kV power transmission and distribution installations. In: China international conference on electricity distribution

4. Kumar P, Singh Y (2010) A software reliability growth model for three-tier client server system. Int J Comput Appl 1(13):9–16
5. Yamada S, Ohtera H (1990) Software reliability growth models for testing-effort control. Eur J Oper Res 46:343–349
6. Crow LH (2010) The extended continuous evaluation reliability growth model. In: Proceedings of the annual reliability and maintainability symposium (RAMS), 2010
7. Krasich M (2012) Power law model, correct application in reliability growth do the cumulative times indeed always add up?. In: Proceedings of the annual reliability and maintainability symposium (RAMS), 2012
8. Strunz R, Herrmann JW (2012) Planning, tracking, and projecting reliability growth a bayesian approach. In: Proceedings of the annual reliability and maintainability symposium (RAMS), 2012
9. Hoyland A, Rausand M (1994) System reliability theory models and statistical methods. Wiley-Interscience Publication, New York
10. Abernethy RB (2000) The new weibull handbook, 4th edn. Robert B Abernethy, Florida

Design for Probe-Type Fault Injector and Application Study of PHM Case

Jun-you Shi, Xiao-tian Wang and Hong-tao Liu

Abstract Prognostics and health management (PHM) is a highlight of current research in modern industrial equipment fields. Aiming at the requirement of verification of PHM system's fault diagnosis capability, a fault injection device is designed and it consists of the digital control unit, fault injection unit, failure mode switching unit, channel switch unit, a status indication unit, data acquisition unit, communication unit and control software unit. The procedures of fault injection experiment are expounded in this chapter. Then a control computer system is applied to a fault injection experiment based on the probe-type fault injector, and the capabilities of fault diagnosis of PHM system are verified. Experiment results show that the probe-type fault injector can be effectively used to verify the PHM fault diagnosis capability of electronic systems and it is non-destructive and easy for operation.

Foundation Item: Major State Basic Research Development Program of China (973 Program) (No.61316705); Basic Research Program (No.A2120110004); Pre-research Foundation (No.51319040301).

J. Shi · X. Wang (✉) · H. Liu
School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China
e-mail: 1035934331@qq.com

J. Shi
e-mail: shijy016@163.com

H. Liu
e-mail: Liuht_buaa@163.com

J. Shi
Science & Technology on Reliability & Environmental Engineering Laboratory, Beihang University, Beijing 100191, China

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_109

1 Introduction

Prognostics and health management (PHM) is becoming the key technology in the research and development of complex weapon equipment system [1]. With the intensive study and extensive application of the PHM technology, it has great meaning to validate and evaluate the PHM system's design capability and it has become a highlight of current research in PHM fields [2, 3]. PHM consists of fault diagnosis, fault prognostics and health management [4] and thus the validation and evaluation of PHM system's design capability includes the verification of fault diagnosis, the validation of fault prognostics and the validation of health management [5].

In this chapter, it mainly involved fault diagnosis capability's validation of PHM system. At present, there are three main approaches to validate and evaluate the PHM system's fault diagnosis capability: the method of basing analysis and valuation, the method based on simulation (including entire system simulation and semi-physical simulation) and the method based on fault injection experiment [6]. Compared with other methods, the method based on fault injection experiment has the advantages of flexible, convenient, low-cost and effective and it is playing a more and more important role in this field.

The method based on fault injection experiment refers to validating and evaluating the PHM system's fault diagnosis capability by injecting fault on the target system or equivalent system to observe and analyze PHM system's running status [7]. There are many different ways in the concrete implementation of the fault injection experiment, including simulation, hardware, software, heavy ion radiation and laser irradiation. In the verification of electronic system's PHM fault diagnosis capability, the hardware implementation is more rapid and effective, but it faces the problems of transient fault unable to be injected, fault injection equipment difficultly inserting target system and easy to cause critical damage on target system [8, 9]. In this background, we design a probe-type fault injector and it is applied to validate the PHM system's fault diagnosis capability of an aircraft's control computer system.

2 The Design of Probe-Type Fault Injector

2.1 The Structure of Probe-Type Fault Injector

The structure of probe-type fault injector is shown in the Fig. 1 and the picture of the probe-type fault injector is shown in the Fig. 2. In the following, the function of each module will be described.

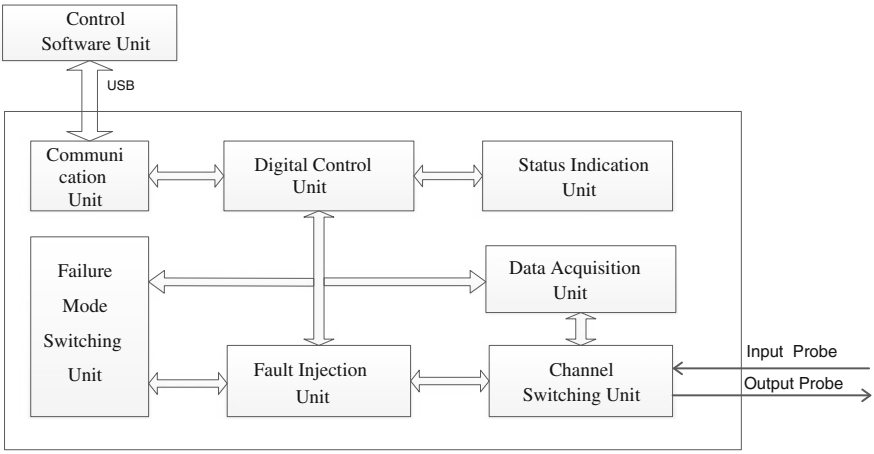


Fig. 1 The structure of Probe-type fault injector

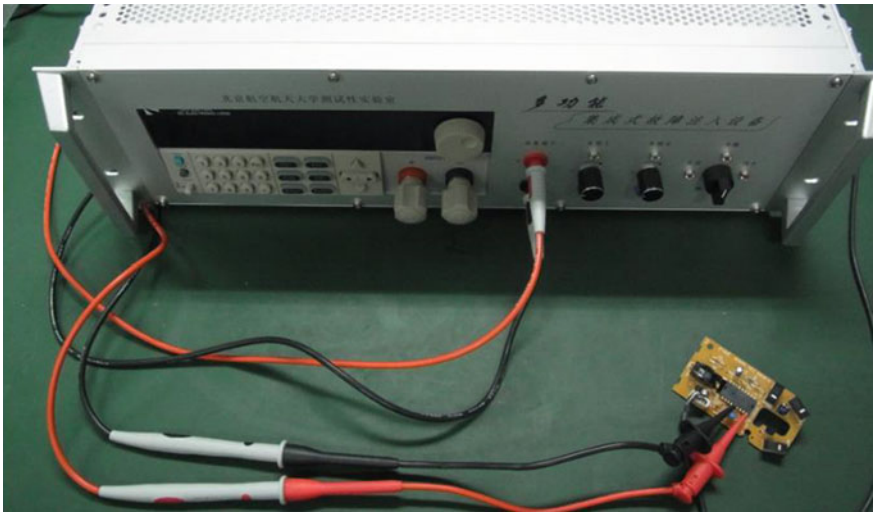


Fig. 2 The Probe-type fault injector

- Digital Control Unit
  The digital control unit is the heart of the probe injector and it can control fault injection channel on–off, failure mode's switch, indication unit's status indicating, and keep fault injector communicating with the control software.
- Fault Injection Unit
  In this unit, it can make kinds of failure modes, such as the stuck at 0 fault, stuck at 1 fault and broken fault by changing the signal from the input probe.
- Failure Mode Switching Unit

According to the needs of fault injection experiment, it can switch the fault injection mode by the failure mode switching unit.

- Channel Switching Unit
  The probe-type fault injector has 16 channels and it can control fault injection channel on–off by the Channel Switching Unit. On the other hand, it provides an access to the fault injector for the input probe and output probe.
- Data Acquisition Unit
  It can collect the signal in the input and output channel and make sure the signal can be conveniently observed and recorded in the fault injection process by this unit.
- Status Indication Unit
  The function of this unit is displaying some basic information about the fault injection, such as the fault time, fault injection mode and fault injection channel, etc.
- Communication Unit
  The main function of this unit is to achieve reliable communication between the fault injector and the control computer.
- Control Software Unit
  The control software is located in the control computer and it can choose the fault injection channel and fault injection mode and control the time and cycle of fault injection by this unit.

2.2 The Main Module Design of the Fault Injector

2.2.1 The Fault Injection Unit

The probe-type fault injector can inject kinds of failure modes in different situations and the Table 1 show the main information of failure mode that can be injected by the probe-type fault injector.

Owing to space constraints, this chapter only elaborates the principle of injecting the Stuck at 0 fault and Stuck at 1 fault. The schematic diagram of the Stuck at 0 fault and Stuck at 1 fault is shown in the Fig. 3.

The heart of Stuck at 0 fault and Stuck at 1 fault module is adjustable high-precision resistance unit and it can adjust the voltage of the input signal in the fault

Table 1 Information of failure mode

Applicable scope	Electronic system
Number of channels	16
Failure modes	Stuck at 0 fault, stuck at 1 fault, broken fault, short fault, inversion fault
Time of duration	Transient fault, continuous fault, intermittent fault

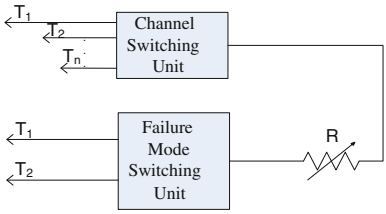


Fig. 3 The schematic diagram of the stuck at 0 fault and stuck at 1 fault

injection experiment according the need of verification experiment. At the begging of fault injection experiment, we should connect the TH probe with the chip's power pin and connect the TL probe with the chip's ground pin. In The Next, we need to connect the Ti (i = 1, 2…n) probe with chip's pin, which need inject fault according to verification experiment and fix the TH, TL and Ti (i = 1, 2…n) probe on the circuit board by insulating cement which can't damage the circuit board in this way. After replacing the circuit board in target electronic system, we should power-on target electronic system and adjust high-precision resistance unit to inject Stuck at 0 fault or Stuck at 1 fault.

2.2.2 Channel Switching Unit

Channel switching unit consists of input and output probe and the relay module. The main function of input and output probe make sure fault injector can reliably connect the target system without serious damage target system. According to the characteristic of electronic system, we designed two types of probe: the needle-like probe and the probe with clip connector. The picture of the two kinds of probe is shown in the Fig. 4.

The relay module consists of sixteen isolated relays and drive circuits and it can control sixteen channels on–off according to the signal from the digital control unit.

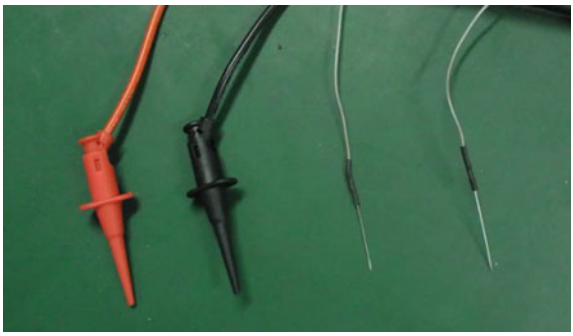


Fig. 4 The picture of the probe

Since the output signal from the digital control unit is unable to drive the relay, it needs the adjunct circuit to drive the relay and we use the Darlington push-pull circuit to drive the relay. After the signal from the digital unit is amplified in the Darlington push-pull circuit, it can control relay on–off and thus control fault injection channel on–off.

2.2.3 Failure Mode Switching Unit

The failure mode switching unit consists of a series of multiple-way switch analogue switch and we choose CD4052BC as multiple-way switch to control the switch of failure mode in the fault injector. The CD4052BC is digitally controlled analogue switch having low "ON" impedance and very low "OFF" leakage currents and it is a differential 4-channel multiplexer having two binary control inputs, A and B, and an inhibit input. The two binary input signals select 1 or 4 pairs of channels to be turned on and connect the differential analogue inputs to the differential outputs. According the need of the verification experiment, it can switch different failure modes by the CD4052BC.

2.2.4 Control Software Unit

Man-machine interface is an important component of the fault injector and we use Visual C++ to develop control software unit in WINDOWS platform. Before the fault injection experiment, the control software assigns an address for the fault injector by transferring API function to ensure the injector can be recognized. To ensure the probe-type fault injector is no hardware fault, it should run self-checking program. After the Power-On Self-Test, the control software unit can choose the fault injection channel and fault injection mode and control the time and cycle of fault injection according the need of the verification experiment. The software flowchart of the control software is shown in the Fig. 5.

3 The Procedure of Fault Injection Experiment

We should analyse the target system to determine the failure mode, the number of how many fault should be injected and the position of where the fault should be injected before the fault injection experiment. In the next, we should connect the target system and the probe-type fault injector and ensure that the communication is

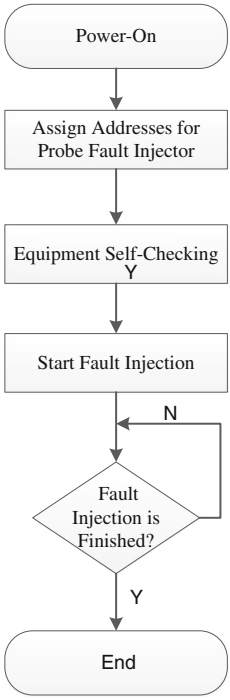


Fig. 5 Control software flow-chart

well between the control computer and the probe-type fault injector. After the fault injection system is constructed, we can inject the failure mode according the need of the fault injection experiment and should make detailed records about the experiment. In the last, we should analyse the experiment data and write the verification report about the PHM fault diagnosis capability. The detailed steps of the fault injection experiment are shown in the Fig. 6.

4 Application Study of PHM Case

After the probe-type fault injector is designed, we construct a verification test on the PHM system of an aircraft's control computer. In electronic systems, fault diagnosis mainly refers to detect and isolate the fault by using the information from the Built-In Test System (BITS). Therefore, it is one of the most important sections to validate the validity of the Built-In Test (BIT) when we verify the PHM fault diagnosis capability of electronic systems.

After the analysis of the control computer's PHM system, we determine to inject fifty faults in this experiment. The PHM system detects fifty faults and isolates forty

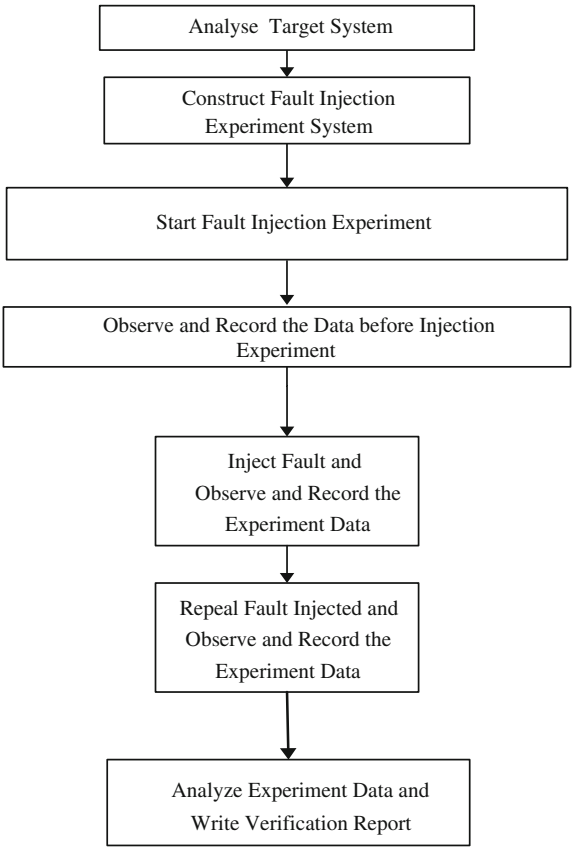


Fig. 6 The steps of fault injection experiment

seven faults. It can be obtained the fault detection rate is 100 % and the fault isolation rate is 94 % after calculation. We successfully validate the PHM system's fault diagnosis capability of the control computer and it has a positive effect on the follow-up work of the PHM system's verification. This chapter

selects four typical signal diagrams which indicate the change of the signal before and after the fault is injected to the he control computer system. Figure 7 shows the signal change after the stuck at 1 fault is continuously injected. Figure 8 shows the signal change after the stuck at 1 fault is intermittently injected. Figure 9 shows the signal change after the stuck at 0 fault is continuously injected. Figure 10 shows the signal change after the stuck at 0 fault is intermittently injected.

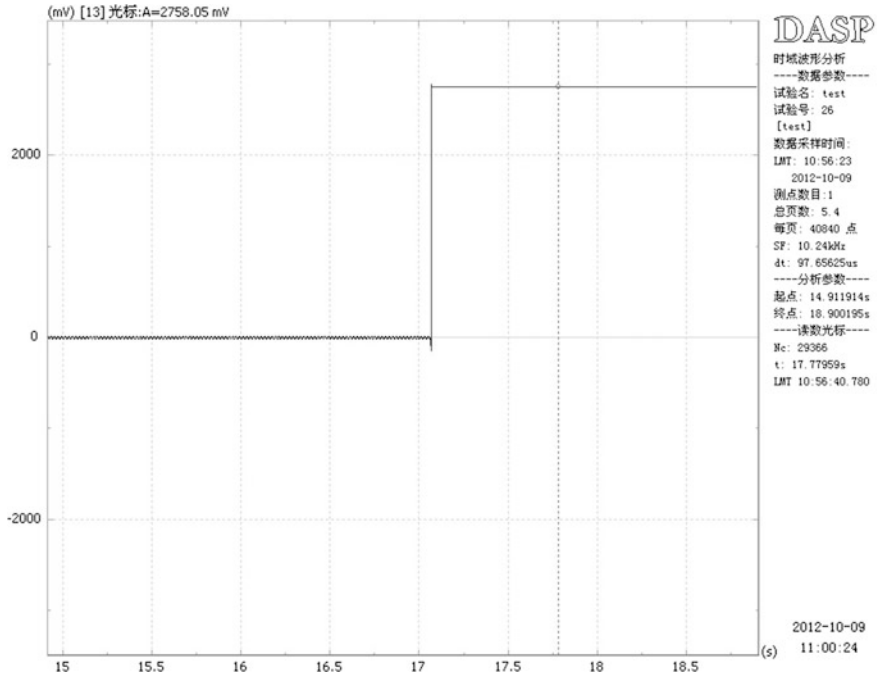


Fig. 7 Continuously inject stuck at 1 fault

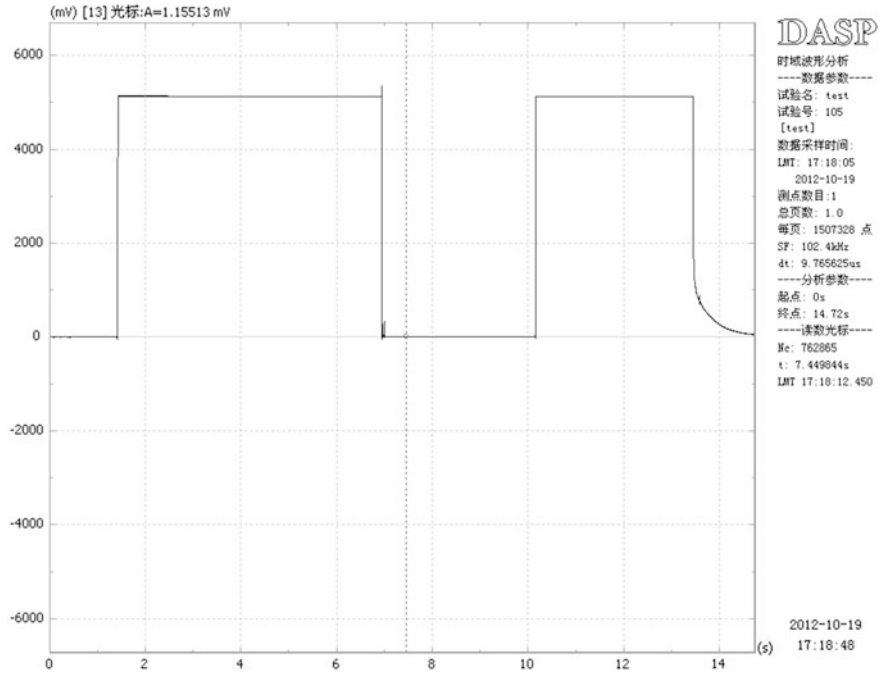


Fig. 8 Intermittently inject stuck at 1 fault

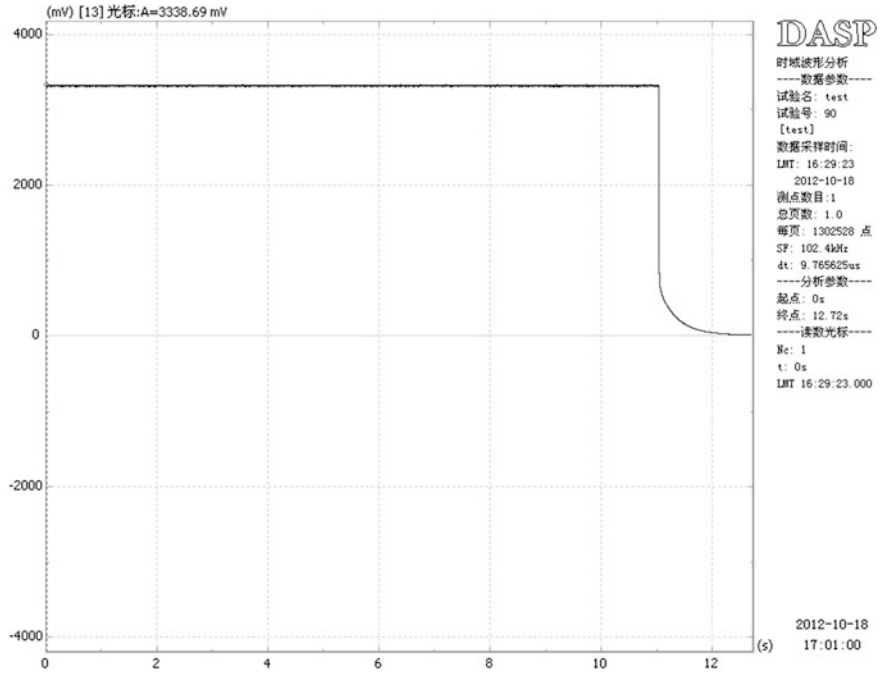


Fig. 9 Continuously inject stuck at 0 fault

5 Conclusion

It is an important part to research fault injector which is applicable for electronic equipment in the work of verifying the PHM fault diagnosis capability of electronic systems. Aiming at the requirement of verification of PHM system's fault diagnosis capability and combining the characteristics of generality, scalability and security the instrument and equipment should have, we designed a probe-type fault injector, which can inject Stuck at 0 fault, Stuck at 1 fault, Broken fault, Short fault and Inversion fault, etc. To verify the effect and availability of the probe-type fault injector, we carried out a fault injection experiment. Through the fault injection experiment, we verify the fault injection capability of the probe-type fault injector. On the other hand, we also find the probe-type fault injector has some drawbacks, such as the fault injection type is not enough, the data acquisition accuracy of data acquisition unit is not precise and the application scope is limited, etc. Therefore, there are still some areas that need us to improve the performance of the probe-type fault injector.

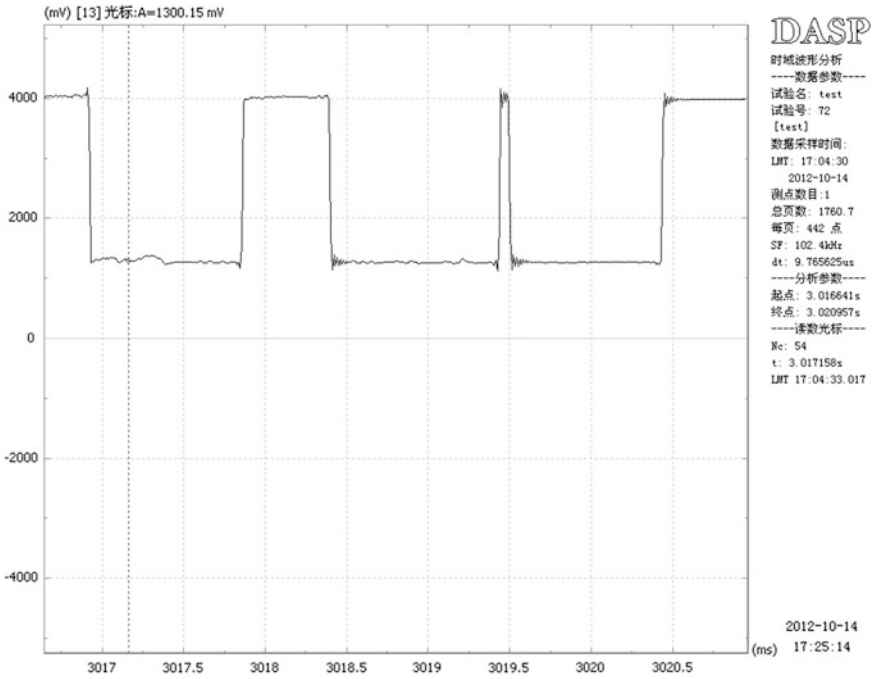


Fig. 10 Intermittently inject stuck at 0 fault

References

1. Bo J, Zhou Y, Jie ZH et al (2011) Review on validation and verification methods of PHM system. Comput Eng Appl 47(21):23–27
2. Junchao S, Jianying W, Xiaozong Y (2000) The present situation for research of fault injector methodology and tools. J Astronaut 22(1):99–104
3. Jun-you SH, Chao J, Hai-wei L (2012) Analysis of testability verification technology and application status. Measur Control Technol 31(5):29–33
4. Jun-you SH, Zheng L, Liu L et al (2007) Design and implementation of automatic control fault insertion equipment. Acta Aeronautica et Astronautica Sinica 28(3):556–560
5. Wang SH, Yu-feng S, Zi-li W et al (2009) A study of PHM system and its fault forecasting model. Fire Control Command Control 34(10):29–35
6. Zhi-peng W, Chen L, Zi-li W et al (2011) Design of PHM demonstration and verification platform. J Nanjing Univ Sci Technol 178:250–255
7. Sheng-kui Z, MichaeI GP, Ji W (2005) Status and perspectives of prognostics and health management technologies. Acta Aeronautica et Astronautica Sinica 26(5):626–632
8. Xiao-min ZH, Zheng-jun ZH, Jie-zhong M (2009) Design and implementation of fault injector based on CPLD. Comput Eng Des 30(17):3921–3924
9. Bao-zhen ZH (2008) Evolution and application of PHM technologies. Measur Control Technol 27(2):5–7

A Method of Establishing the Dependency Integrated Matrix Based on Diagonally Dominant Fuzzy Transitive Matrix

Tong Zhang, Jun-You Shi and Yin-Yin Peng

Abstract For the purpose of detecting the state of systems or equipment, and isolating the internal faults in the field of testability, a method of establishing the dependency integrated matrix based on diagonally dominant fuzzy transitive matrix is proposed. The principle of dependency integrated matrix is introduced. On the basis of fuzzy transitive modeling towards the faults and the signs of systems or equipment under detection, the fuzzy transitive matrix is achieved [1]. Furthermore, according to the diagonally dominant matrix, the zero rows (rows with all zero elements) of dependency matrix are settled, and the dependency integrated matrix applied to fault detection is proposed based on the transitive relationship between the faults and the symptoms with the maximum degree of membership [2, 3]. The flow of establishment for dependency integrated model is described in detail. The statistics of a certain circuit are taken as an example for application, which demonstrates this method is feasible and effective.

Keywords Testability · Dependency integrated matrix · Diagonally dominant · Fuzzy transitive matrix

T. Zhang (✉) · J.-Y. Shi (✉)
School of Reliability and Systems Engineering, Beihang University, Beijing, China
e-mail: babyformula@me.com

J.-Y. Shi
e-mail: jy.shi@yahoo.com

Y.-Y. Peng (✉)
Science and Technology on Reliability and Environmental Engineering Laboratory, Beihang University, Beijing, China
e-mail: 13401155359@189.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_110

1 Introduction

Diagonally dominant matrix is a widely used class of matrices, which attaches great importance to the application in information theory, systems theory, modern economics, networks, algorithms, program design, and so forth. In the field of testability, It exercises a far-reaching influence on application to analyze diagonally dominant fuzzy transitive matrices, and establish the dependency integrated matrices that present the correlations between faults and tests on the basis of analysis.

The theory of dependency correlation has played an indispensable role in the technique of testability configuration and analysis [4]. It proposes the dependency correlations between faults and tests by means of testability model and dependency integrated matrix, which bases on single fault assumption. In accordance with dependency correlations, the ordination for tests can be achieved.

The fuzzy transitive matrix indicates the probability relations between symptoms and causes of faults, while the dependency integrated matrix proposes the certain logic correlations between failure modes and tests. Only when the symptoms and faults refer to certain tests and failure modes does fuzzy transitive matrix interrelate to the corresponding dependency integrated matrix. Meanwhile, both of them are interconverted in the sense of mathematics, in which the fuzzy transitive matrix R^T is available to be obtained through engineering data among tests, not compatible for applying directly because of the variety data, while the dependency integrated matrix D fulfills the condition that quick and intuitive. To sum up, the latter would be an added asset in engineering projects.

2 Related Definitions

The complete process of testability modeling contains collection, collation, digestion and modelling. It is complicated as short of the unified method of data collection and processing. During the actual modeling process, there are generally related definitions in the field of testability as follows [5]:

Definition 1 Diagonally Dominant Fuzzy Transitive Matrix

Let (r_{ij}) donate an $m \times n$ matrix, $r_{ij} \in \mathbf{R}^{m \times n}$, and $m \geq n$.

The matrix (r_{ij}) is diagonally dominant fuzzy transitive matrix if

$$|r_{ii}| \geq \sum_{j=1, j \neq i}^{n} |r_{ij}|, \ i = 1, 2, \ldots, n \tag{1}$$

In (1): r_{ij} is the probability that i-th fault Y_i causes j-th symptom X_j. And $|r_{ij}|$ is the absolute value of r_{ij}.

Definition 2 Dependency Integrated Matrix

Let (d_{ij}) donate an $m \times n$ matrix.

The matrix (d_{ij}) is dependency integrated matrix if

$$d_{ij} \in \{0, \quad 1\},\ i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n \tag{2}$$

In (2): d_{ij} presents the correlation between i-th failure mode F_i and j-th test T_j.

$$d_{ij} = \begin{cases} 0, & T_j \text{ cannot detect } F_i, \text{ when } T_j \text{ observethe logic value } (T_j \text{ and } F_i \text{ areunrelated}) \\ 1, & T_j \text{ can detect } F_i, \text{ when } T_j \text{ observe the logic value } (T_j \text{ and } F_i \text{ are related}) \end{cases}$$

Definition 3 Let Y be the set of all existing faults in a UUT (Unit Under Test), then

$$Y = \{Y_1, Y_2, \ldots, Y_m\} \tag{3}$$

In (3): m is the amount of variety faults.

Definition 4 And let all symptoms that the faults cause be X, shown as:

$$Y = \{Y_1, Y_2, \ldots, Y_n\} \tag{4}$$

In (4): n is the amount of symptoms.

Definition 5 As the faults correspond to the symptoms, the matrix R based on the X and Y can be achieved according to the fuzzy logic, shown as:

$$R_{n\times m} = \begin{array}{c} \\ X_1 \\ X_2 \\ \vdots \\ X_n \end{array} \begin{array}{c} \begin{array}{cccc} Y_1 & Y_2 & \cdots & Y_m \end{array} \\ \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{bmatrix} \end{array} \tag{5}$$

In (5): X_j is the j-th fault, Y_i is the i-th symptom, and r_{ij} is the grade of membership value that i-th symptom corresponds to j-th fault, $0 \le r_{ij} \le 1$, $1 \le i \le n,\ 1 \le j \le m$.

Definition 6 Same as R^T, the fuzzy transitive matrix, which is the transpose of the matrix R and also called fuzzy diagnosis matrix in the field of fault diagnosis, as below:

$$R^T_{m\times n} = \begin{array}{c} \\ Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{array} \begin{array}{c} \begin{array}{cccc} X_1 & X_2 & \cdots & X_n \end{array} \\ \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} \end{array} \tag{6}$$

Dependency integrated matrix presents the logic correlations between failure modes and tests. In a UUT comprised of m units, Let F_i be the faults of variety units, and d_{ij} indicates the logic correlation with the n corresponding test points.

Definition 7 The dependency integrated matrix of the UUT is proposed as below:

$$D_{m\times n} = \begin{array}{c} \\ F_1 \\ F_2 \\ \vdots \\ F_m \end{array} \begin{array}{c} \begin{array}{cccc} T_1 & T_2 & \cdots & T_n \end{array} \\ \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{bmatrix} \end{array} \tag{7}$$

In (7): $F_i = [d_{i1}, d_{i2}\ldots d_{in}]$ is the dependence of i-th failure mode and the test set $T_j(j = 1, 2, \ldots, n)$. The $T_j = [d_{1j}\ d_{2j}\ldots d_{mj}]^T$ is the j-th dependency of the test and the failure mode $F_i(i = 1, 2, \ldots, m)$. And the value of d_{ij} is as below:

$$d_{ij} = \begin{cases} 0, \ T_j \text{ cannot detect } F_i, \text{ when } T_j \text{ observethe logic value } (T_j \text{ and } F_i \text{ areunrelated}) \\ 1, \ T_j \text{ can detect } F_i, \text{ when } T_j \text{ observe the logic value } (T_j \text{ and } F_i \text{ are related}) \end{cases} \tag{8}$$

3 Algorithms and Flowcharts

The objective method proposes to establish the dependency integrated matrix based on diagonally dominant fuzzy transitive matrix, and aims at further tests though the obtained matrix [6, 7]. The entire flow of this method is demonstrated in Fig. 1.

Specifically steps of the process as follows:

i. Apply fuzzy transitive modeling to faults and symptoms of the system under test, and achieve the fuzzy transitive matrix $R^T = \left(r_{ij}\right),\ r_{ij} \in \mathbf{R}^{m\times n},\ m \geq n,$. The element r_{ij} indicates the probability that i-th fault Y_i causes j-th symptom X_j [8]. During the project application, there is no condition that the signs of fault appear while no fault occurred, i.e. it is impossible that $n > m$ in this method. Then judge the fuzzy transitive matrix R^T if diagonally dominant, see details in Definition 1.
ii. Transform the matrix R^T into diagonally dominant matrix. The process of arranging fuzzy transitive matrix R^T into diagonally dominant matrix $R^{T\prime}$ is shown in Fig. 2.

    Specifically steps of the process as follows:

    (1) Begin with the i-th row of the fuzzy transitive matrix R^T, initial i = 1;
    (2) Let j = 1, variable k = 0;

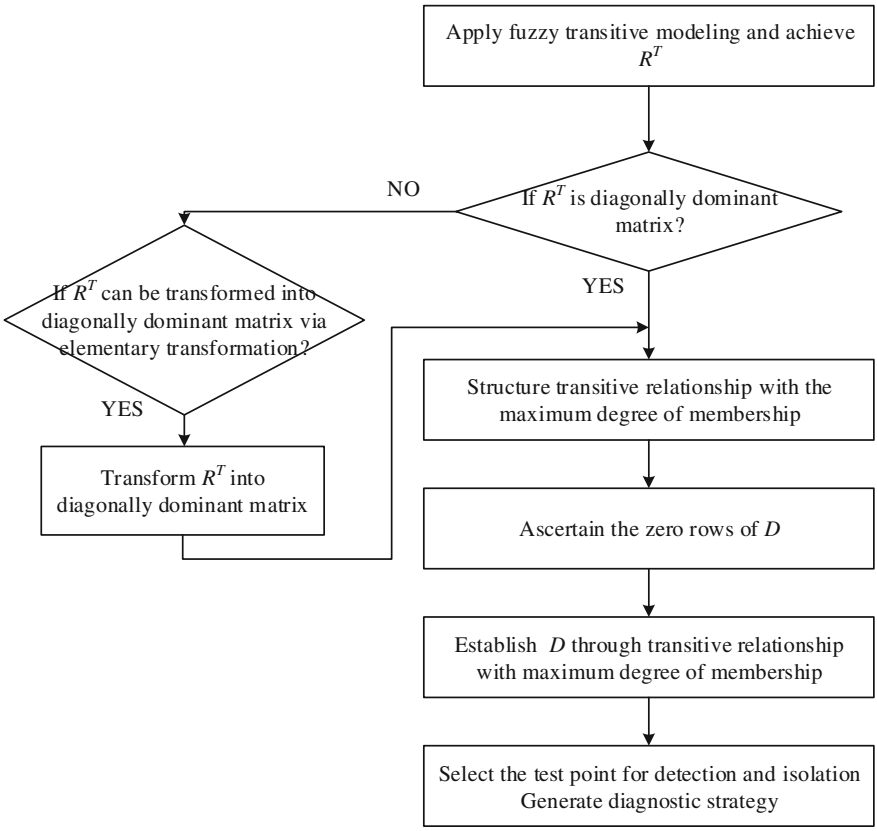


Fig. 1 The process of establishing the dependency integrated matrix

(3) Select the j-th element, judge if $|r_{ij}|$ is greater than the sum the absolute value of the rest elements in the j-th rank of the fuzzy transitive matrix R^T, if yes, go to (4); or go to (5);
(4) Record $|r_{ij}|$, and update $k = k + 1$;
(5) Update $j = j + 1$, and judge if j is greater than n, if yes, go to (6); or go to (3);
(6) Judge if $k > 1$, if yes, the fuzzy transitive matrix cannot be transformed into diagonally dominant fuzzy transitive matrix, break; or go to (7);
(7) Update $i = i + 1$, and judge if $i > m$, if yes, go to (8); or go to (1)
(8) Towards the previous $|r_{ij}|$, exchange i-th and j-th rank of the fuzzy transitive matrix R^T. The exchanging information is obtained, and the transformation of the fuzzy transitive matrix into diagonally dominant fuzzy transitive matrix is done, break.

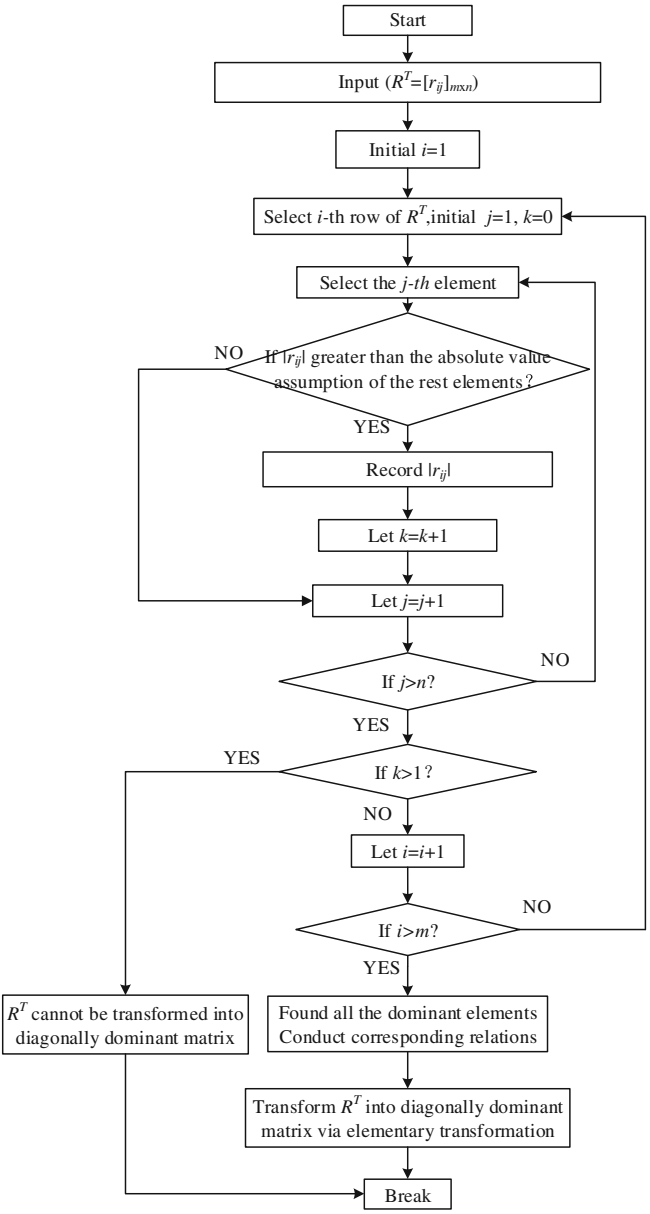


Fig. 2 The process of arranging fuzzy transitive matrix R^T into diagonally dominant matrix $R^{T\prime}$

Sort out the diagonally dominant fuzzy transitive matrix $R^{T\prime}$and the correlations between faults and symptoms.

iii. In accordance with the diagonally dominant fuzzy transitive matrix $R^{T\prime}$, structure the transitive relationship between the faults and the symptoms with the maximum degree of membership. The fault Y_i and the symptom X_i that r_{ii} corresponds indicates the maximum probability that Y_i causes X_i, and X_i appears. Set up the corresponding relation $X_i \rightarrow Y_i$, and "→" coveys the corresponding relation [9–12].

iv. Let D be the requested dependency integrated matrix, $D = \left(d_{ij}\right), d_{ij} \in \mathbf{R}^{m\times n}$. The element d_{ij} indicates the logic correlation between i-th failure mode F_i and j-th test T_j. Ascertain the zero rows of the corresponding dependency matrix for tests according to the corresponding relations: the p-th row in matrix D is zero, $n+1 \leq p \leq m$.

v. Establish the dependency integrated matrix D through the transitive relationship between the faults and the symptoms with the maximum degree of membership from the fuzzy transitive matrix R^T. The process of setting up zero rows of dependency integrated matrix D is shown in Fig. 3.

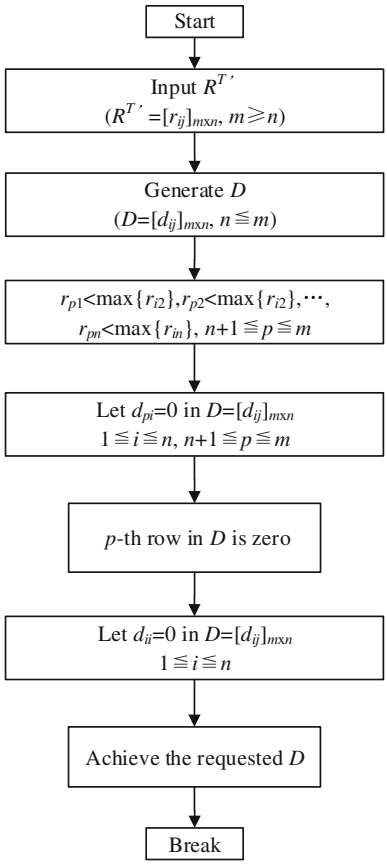


Fig. 3 The process of setting up zero rows of dependency integrated matrix D

vi. Let $d_{ii} = 1$, $d_{ij} = 0$, $(i \neq j)$, and the requested dependency integrated matrix D is achieved as below:

$$D = \begin{array}{c} \\ F_1 \\ \vdots \\ F_n \\ F_{n+1} \\ \vdots \\ F_m \end{array} \begin{array}{c} \begin{array}{ccc} T_1 & \cdots & T_n \end{array} \\ \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \end{array}$$

vi. Select the test point for detection and the test point for isolation, generate the diagnostic strategy, so as to further detection and localization, in accordance with the acquired dependency integrated matrix D [12].

4 Application

For the signal disposal circuit of an avionics module, the fuzzy transitive matrix R^T is obtained by modeling to the faults and the symptoms as below:

$$R^T = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \end{array} \begin{array}{c} \begin{array}{ccccc} X_1 & X_2 & X_3 & X_4 & X_5 \end{array} \\ \begin{bmatrix} 0.018 & 0.546 & 0.027 & 0.081 & 0.028 \\ 0.546 & 0.046 & 0.046 & 0.059 & 0.046 \\ 0.084 & 0.052 & 0.574 & 0.062 & 0.074 \\ 0.04 & 0.066 & 0.048 & 0.574 & 0.073 \\ 0.066 & 0.079 & 0.076 & 0.019 & 0.68 \\ 0.09 & 0.05 & 0.03 & 0.03 & 0.07 \\ 0.06 & 0.075 & 0.075 & 0.092 & 0.098 \end{bmatrix} \end{array}$$

$Y_1 \sim Y_7$ means the 1st $\sim$ 7th fault, $X_1 \sim X_5$ means symptoms that the 1st $\sim$ 7th fault causes, r_{ij} indicates the probability that i-th fault Y_i causes j-th symptom X_j.

The fuzzy transitive matrix R^T is diagonally dominant matrix judging from Definition 1. Afterwards, transform the matrix R^T into diagonally dominant matrix. In the fuzzy transitive matrix R^T, exchanging the rank X_i and Y_j, does not interrupt the correlation between the certain fault and symptom. Therefore, R^T can be considered diagonally dominant if it satisfy Definition 1 after limited time exchange.

Sort out the diagonally dominant fuzzy transitive matrix $R^{T\prime}$.It can be noticed that $r_{12} = 0.546$ in 1st row; $r_{21} = 0.546$ in 2nd row; $r_{33} = 0.574$ in 3rd row; $r_{44} = 0.574$ in 4th row; $r_{55} = 0.68$ in 5th row, which meets the Definition 1.

Exchange the 2nd and the 1st rank, $R^{T\prime}$ is obtained as below:

$$R^{T\prime} = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \end{array} \begin{array}{c} \begin{array}{ccccc} X_1 & X_2 & X_3 & X_4 & X_5 \end{array} \\ \begin{bmatrix} 0.546 & 0.018 & 0.027 & 0.081 & 0.028 \\ 0.046 & 0.546 & 0.046 & 0.059 & 0.046 \\ 0.052 & 0.084 & 0.574 & 0.062 & 0.074 \\ 0.066 & 0.04 & 0.048 & 0.574 & 0.073 \\ 0.079 & 0.066 & 0.076 & 0.019 & 0.68 \\ 0.05 & 0.09 & 0.03 & 0.03 & 0.07 \\ 0.075 & 0.06 & 0.075 & 0.092 & 0.098 \end{bmatrix} \end{array}$$

Verify the $R^{T\prime}$ satisfy Definition 1.

Set up the corresponding relation $X_i \rightarrow Y_i$, and achieve $X_1 \rightarrow Y_1, X_2 \rightarrow Y_2, \ldots, X_5 \rightarrow Y_5$. Therefore, p-th row is zero in the dependency integrated matrix D, $n+1 \leq p \leq m$, i.e. $d_{p1} = d_{p2} = \cdots = d_{pn} = 0$.

Generate the requested dependency integrated matrix in accordance with established corresponding relations and the zero rows as below:

$$D = \begin{array}{c} \\ F_1 \\ F_2 \\ F_3 \\ \vdots \\ F_n \\ F_{n+1} \\ \vdots \\ F_m \end{array} \begin{array}{c} \begin{array}{ccccc} T_1 & T_2 & \cdots & T_{n-1} & T_n \end{array} \\ \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ 0 & \vdots & \ddots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \end{array}$$

Specifically shown as below:

$$D = \begin{array}{c} \\ F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \\ F_6 \\ F_7 \end{array} \begin{array}{c} \begin{array}{ccccc} T_1 & T_2 & T_3 & T_4 & T_5 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Based on the obtained dependency integrated matrix D, select the detecting test point and the isolating test point, generate the diagnostic strategy, consisting of sequence for detection, fault isolation and fault diagnostic tree [13–15], so as to further detection and localization, in accordance with the acquired dependency integrated matrix D.

In the generated matrix D, the weighted value of fault detection of each test $W_{FD} = 1$, select the test T_3 which can simplify the process. And the weighted value of fault isolation of each test $W_{FI} = 5$, therefore, select the test T_3 that has been proposed as the test point for detection [16].

5 Conclusion

This paper provides a standardized method of establishing the dependency integrated matrix based on diagonally dominant fuzzy transitive matrix, aiming at generating the dependency integrated matrix in accordance with the diagonally dominant fuzzy transitive matrix, and conducting fault detection through the achieved matrix.

The method supplies a gap of generating the dependency integrated matrix through known fuzzy transitive matrix, helps generate the transitive correlations between faults and symptoms intuitively and feasibly, and achieve the requested dependency integrated matrix for fault detection and isolation. Therefore, further operation of diagnosis and fault localization can be conducted.

References

1. Choi GS, Iyer RK, Saleh R, Carreno V (1991, January) A fault behavior model for an avionic microprocessor: A case study. In Dependable computing for critical applications. Springer, Vienna, pp 177–195
2. Goode PV, Chow MY (1993, November) Neural/fuzzy systems for incipient fault detection in induction motors. In industrial electronics, control, and instrumentation, 1993. proceedings of the IECON'93., international conference on (pp 332–337). IEEE
3. Isermann R (1984) Process fault detection based on modeling and estimation methods—a survey. Automatica 20(4):387–404
4. Shi JY (2011) Testability design analysis and verification. In National Defence Industry Press, pp 90–119
5. Kahraman C, Ertay T, Büyüközkan G (2006) A fuzzy optimization model for QFD planning process using analytic network approach. Eur J Oper Res 171(2):390–411
6. Liu H, Zhao W, Zhang J, Wang L, Ma X, Zhao F (2011, May) Modal analysis of machine tools during working process by matrix perturbation method. In: Assembly and manufacturing (ISAM), 2011 IEEE international symposium on (pp 1–4). IEEE
7. Liu WY (1993) The fuzzy functional dependency on the basis of the semantic distance. Fuzzy Sets Syst 59(2):173–179
8. Yang P, Qiu J, Liu GJ (2008) Research on extended dependency model-based testability analysis. Systems Engineering and Electronics, February 2008, pp 371–374
9. Patton RJ, Hou M (1998) Technical communique: design of fault detection and isolation observers: a matrix pencil approach. Automatica (J IFAC) 34(9):1135–1140
10. Priya RD, Sujitha S, Jeevitha S (2013) Estimation of age group most affected by various reasons for the drop outs of the students in coimbatore by using fuzzy matrix. Int J Math Arch (IJMA) ISSN 2229-5046, 4(4)

11. Shi JY, Lin XG, Shi M (2012). A key metric and its calculation models for a continuous diagnosis capability base dependency matrix. Metrol Meas Syst 19(3):509–520
12. Tian Z, Shi JY (2013). System testability design, analysis and verification. Beijing, China
13. Shi JY, Zhang X, Zou TG (2011) Application of multi-signal modeling and diagnosis strategy design technology. Syst Eng Elect 33(4):811–815
14. Shi J, Lin X, Lv K (2012, May) A method for searching and evaluating diagnosable sequence fault sets of a dependency matrix. In: Prognostics and system health management (PHM), 2012 IEEE conference on (pp 1–7). IEEE
15. Shi J, Zhang T, Wang F (2012, May) A data pre-processing method for testability modeling based on first-order dependency integrated model. In: Prognostics and system health management (PHM), 2012 IEEE Conference on (pp 1–8). IEEE
16. Trampert J, Fichtner A, Ritsema J (2013, April) Estimating resolution in full waveform tomography. In: EGU General Assembly Conference Abstracts (vol 15, p 3098)

Understanding and Evaluating IT Budgets and Funding

Indira Venkatraman and Paul T. Shantapriyan

Abstract The average firm spends over two-thirds of the IT budget on maintaining the present infrastructure and architecture. The IT budget in turn makes up between 30 and 50 % of the capital expenditure of a firm annually. An appreciation of the models would enhance an understanding of funding for new IT initiatives and/or making improved decisions on IT investments and operations. In an era where IT infrastructure is essential, a sound IT portfolio is important. This chapter explores the necessities of maintaining a firm's current IT health, investing in future capacity and improving the competitive position of the firm. This chapter tries to suggest how an organisation can shift from barely maintaining the current IT infrastructure, much to the chagrin of everyone involved, to an industry leading, value generating and sound IT infrastructure. This chapter first examines four models for Budgeting and Funding for Information Technology in a firm (Charge Based, Lord and Master, Revenue Percentage, Absolute Outsourcing). Then it proceeds to develop a speculative framework to assist IT budget and funding. There is evidence that the first four models (Charge-based, Lord and Master, Revenue Percentage and Absolute Outsourcing). Charge-based is a model where the IT unit is treated outright as a service provider and paid for services implemented. In Lord and Master, a firm's IT unit has absolute control over implementation and execution of IT services. Revenue Percentage is where the IT unit compete for budgets in the form of a percentage of the organisation's overall revenue/profits. Absolute Outsourcing is also a service-charge based system; the only difference being that there is no defined IT unit within firm. IT needs are outsourced to an IT firm outside of the organisation and services rendered, charged back to the organisation. The impact an IT department has, whatever model it uses to exist, on organizational activities is un-questionable. The study of these models will help firms gather knowledge as to where they stand and also analyse if they need a change in their IT related financial

I. Venkatraman (✉)
Intelligent Capital, Hobart, Australia
e-mail: indira@intelligentcapital.info

P.T. Shantapriyan
Tasmania school of Business and Economics, University of Tasmania, Hobart, Australia
e-mail: Paul.Shantapriyan@utas.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_111

environment. The proposed models are evaluated on: how each model works, how much is allocated as an IT budget, how the financial relationship of IT with other departments and stake holders functions, and implications on organizational learning. Using the funding and budgeting models as building blocks future research is welcomed on governance, change management as well as operational, day to day practical decision making which has relevance to both academic and practitioners.

Keywords Performance monitoring and management · IT budgets · Funding models

1 Introduction

The budgeting for Information Technology (IT) is viewed from two perspectives: the capital budgeting expenditure and the operating expenditure. Capital expenditure is investment in: infrastructure assets to establish the IT, the purchase of the software, the cost of implementation, training and investment in human resources to leverage the capabilities of IT. Around 46 % of the total IT investment is on infrastructure [1]. The cost of implementing IT from an enterprise perspective has grown, with challenges that the time and budget are often exceeded as well as the value of such IT to the decision maker questioned. Once the IT is established in an enterprise, the operating budget becomes the next subject to debate. The operating expenditure is the recurrent yearly spend on: salaries and wages of IT professionals, maintenance of the software and infrastructure; and licence charges for the software. Within each of these two perspectives are two further expenditures, mandated and discretionary [2]. Mandated expenditures are outlays which are necessary while discretionary outlays are those that can be delayed or avoided completely. As the costs of IT grow, there is increasing pressure for senior management to defer, delay or not invest in further capital expenditure or cull the operating expenditure of IT. Therefore, what should be mandated can be from the view of financially constrained managers, becomes discretionary. Often, the existing charging system for IT can shape the mind set of senior management. One important aspect of investing and budgeting for IT is how savvy the management team is with regard to IT.

2 IT Savvy Firms and Their Funding

Being IT savvy does not require everyone in the firm to learn everything IT, An overall knowledge of the basic principles and their own IT needs should suffice. There are five characteristics for the IT savvy enterprise [1]:

1. IT for internal and external communication. The intensity and frequency of use of IT for internal and external communications shapes how savvy a firm is;
2. Internet use. The use of open internet architectures for building customer, employee engagement, performance management and learning.
3. Digital Transactions. The percentage of digitization undertaken by customers, employees and managers;
4. Companywide skills. The captures the technical and business competencies of the IT human resources.
5. Management involvement. This reflects the commitment by senior management as well as the degree to which IT is diffused throughout the enterprise.

Those enterprises which have a low IT savvy would invest less in infrastructure and strategic information capture and invest more in low risk transactional systems that provide the information for day to day operations [1]. Such decisions would be to reduce the magnitude of the IT expenditure and the exposure to longer term infrastructure assets [1]. The urge to reduce IT costs has seen the use of cloud technology [3] where the use of Infrastructure as a Service (IaaS) can reduce the upfront investment in infrastructure costs (processing, storage and other security resources) of the firm [4].

Having an open mind towards new developments would be good. IT funding models at IT savvy firms have three important components [5];

1. Senior executives establish clear priorities and criteria for IT investments
2. Management develops a transparent process for assessing potential
3. Monitor impacts of prior investment decisions and uses the learning in future

In an era where IT infrastructure is an essential despite what an organisation does for business, a sound IT portfolio is as important as the value of company shares and the nature of people hired to work for the firm. This chapter develops a portfolio based approach to IT investment. How IT money is distributed and used in a firm is based on whether or not the firm is IT savvy or IT ignorant. Various areas of a business that this topic might affect are looked at; for example cost control, learning curves, governance challenges and change management. Finally all this analysis will lead on to the discussion of the speculative framework titled Portfolio Model.

3 Overview of Models

3.1 Charge Based Model

In this scenario the IT department within an organisation behaves like an in-house consultancy or service provider that bills individual departments and the management for providing both soft and hard services in IT. The business units get budget allocated to them for IT expenditures, then IT units charge based on an “internal pricing system” [6].

The percentage of money allocated for IT thus depends on how much each unit demands and/or are willing to spend on IT. Although this model keeps a clear account of what IT services have been requested and provided, it fails to give a big picture of how the entire firm benefits from IT and how much the IT unit really needs to progress, not just give what is requested. A sense of detachment may arise in the morale of the IT unit. They might feel like an external unit and be retreated so by other units. Their rapport might be purely service based and a sense of belonging might be lost.

3.2 Lord and Master Model

In this model the IT unit gets a budget and is allowed to make almost all decisions about all the IT needs and service requirements of the entire firm as a collective. Though it might seem like sensible centralised control one might argue that this system puts too much control in the hands of the IT managers. Conflicts might arise on how IT services are to be provided and even on what type of hardware and/or software is to be provided. Unlike charge-based the people working in the IT unit might feel as part of the entire company. How the IT unit uses the money allocated can be a cause for conflict. Business units within the firm might see the lord and master approach as a takeover by all things IT.

3.3 Revenue Percentage Model

This scenario can take two roads. One, the IT unit/department can have a pre-existing budget and can be eligible for a percentage of the revenue over and about the existing budget. Second, the entire IT budget can be based on a revenue percentage. In this case, after initial set up costs, an IT unit has to keep working at contributing to the revenue increase of a firm to get more funding.

Though such a model has the benefits of meeting the needs of business units operating in a competitive market place, the reward for the IT unit would be limited by the IT non savvy nature of the business unit managers. IF there is no growth in the Strategic Business Units (SBU), this system can demoralise people working for the IT unit if there is not much to work towards. Unless a firms overall IT needs are dynamic and the SBU decision making align with the corporate strategy, there are limits to what an IT unit can do/provide. Unless the funding for all/most units are revenue percentage based, the other units might view this as unfair. Senior Management might find it hard to assess the contributions of IT to the business as tangible evidence for the percentage. Therefore, the separability of the IT co-creation of value can raise questions about investing further in IT for future periods.

3.4 Absolute Outsourcing Model

Some companies do not see the need to have an IT department in the organization at all. In this way, the capital expenditure on hardware, implementation and software costs are avoided as IT is viewed as a support service. At times some firms only need high speed internet and/or video conferencing abilities. In those cases they might find that contracting an IPS service provider is a better deal than having a few extra employees running around setting up something. This also applies to smaller firms who prefer to outsource their Accounting information system (AIS). Application Service Providers (ASP) and Business Process Outsourcing (BPO) are in these lines.

While it has the advantage of experienced outsource agents to handle an organizational IT needs, there can be issues of who owns the data, risk management, and privacy of customer data. The party offering the services is basically a stranger; this can play on the morale and trust levels of employees. In the event an IT budget is supplied to each business unit to afford IT services, there is the possibility of individual units hiring different service providers and private company information being up for grabs. Senior management on the other hand might find it difficult to customise and cost save when time comes to update the IT provisions if the provider is an outsider. They might not intrinsically understand the companies IT needs and alignment.

4 Where to from Here?

One could argue that IT and related fields are just new fads or enhancements that make the process of business more refined and easy to follow. Well it is true that trade, manufacture, accounting, investing, medicine, law, etc., etc. have existed powerfully way before computers were even invented. But one needs to just look around to realise that if we all decided to wind the clock back to a time where computers did not exist, life would be boring indeed and tasks that are a breeze with IT would end up being anvils around our neck.

IT has proceeded far beyond computers, their software and has become an environment, an eco-system if you must that has evolved from the human mind into a tool or prospect that can enhance the way we perform and earn.

So what now? Let us take a portfolio view of IT. A firm can go on treating every IT need as a project or step a bit forward and include IT as a company's portfolio [5]. In the portfolio approach, the IT investment is broken down into four asset classes.

The IT asset classes are;

1. Strategic IT
2. Informational IT
3. Transactional IT
4. IT infrastructure

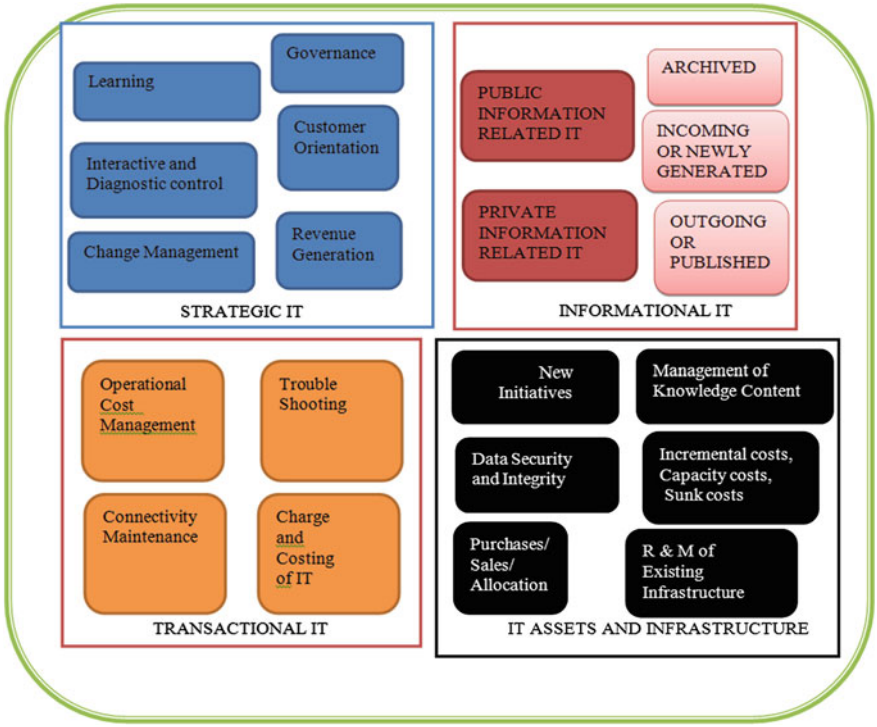


Fig. 1 Portfolio model

The framework is based on the portfolio classes and various other parameters. End of the day a firm has to "Follow the Money"; IT or no IT. So reverse engineering from an investment perspective seemed nice. We are calling it Portfolio Model. The suggested framework is hopefully developed into a sort of box of tools. You know it is there, but you do not have to use every tool every single time. They are there when you need it and are easily put out of sight until needed again. We are taking the four asset classes one step further. We decompose each asset class into key factors that will generate future economic benefits. The following diagram portrays the factors dealt within each portfolio class (Fig. 1).

4.1 Strategic IT

This is basically IT services and processes that assist the strategy, alignment and decision making needs of a firm. Rather than being a top down approach, we prefer to keep this portfolio group as a collection of fields. Like packets of data that make

up a massive cluster of an important strategic tool. Parts of this cluster are (but not limiting to)

- Governance
- Learning
- Interactive and Diagnostic Control
- Change Management
- Customer Orientation
- Revenue Generation.

Corporate governance of IT [7] revolves around the use of IT to enable: strategic decision making, the integration of technical choices, the coordination of resources and structure of IT, the meeting of business unit needs and expectations and the prioritization of investment decision [8].

In appreciating the complexities of strategic investment in IT, accounting can at time lead to myopic decision making. One method commonly used in Discounted Cash Flows of incremental cash flows. A new project would have to generate future cash flows which when discounted would exceed the initial outlay. As investments in IT grow to tens of millions to hundreds of millions, the likelihood is low for any increased cash flows from operations to justify such large outlays. What is not addressed in traditional DCF is that if there is no investment, customers may walk away from the business to competitors. What needs to be discounted is the total shaded area on the right of Fig. 2:

The accuracy of estimating the fall in customers is an imprecise science [9]. However, learning what customers want, how competitors are positioned in the market place and the likely response from the organization is pivotal to generating future revenues. For example, a university that invests in internet based delivery can find future masters applied programs being designed which attract new customers that would not have been possible. This view of options pricing for investing in IT

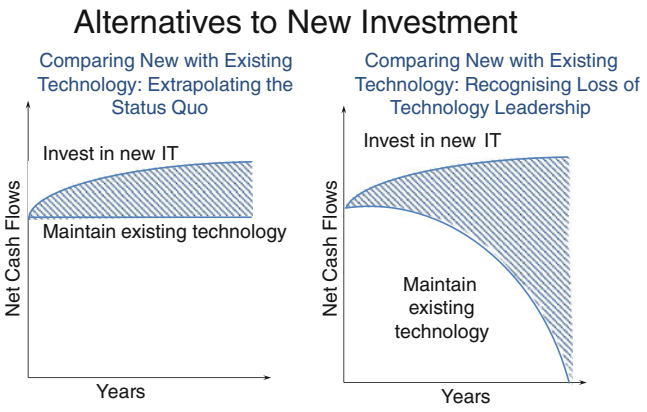


Fig. 2 Alternatives to new investment. (Adapted from Kaplan and Atkinson 1998 [13]; Christensen, Kaufman, Shih 2008 [14])

is gaining credence. We suggest that the future value of the option of IT is dependent on the project teams that use "big data" or business intelligence to have the freedom to analyse data, processes and opportunities to create this value [3]. To facilitate such decision making, certain sets of information would be basis of debate with managers and employees where this interactive use of information [10] would direct and shape the strategic intention of an organization. The ability to manage change as a result of this set of interactive levers of control can enable an organization to optimize the core existing products, expand into adjacent product lines or develop breakthrough products and create new markets [11].

4.2 Informational IT

The main influence that IT had brought about is in the area of information and information handling. Most companies would in their process of establishment and continuity produce both private and public information sets. Informational IT covers the areas of Archiving this Data and maintaining the privacy if the private information by means of security and assurance.

Data/Information generation is not a one of process. It is continuous and ongoing iterative process. As project teams learn and mature in their IT savvy approach to generating value, so does the nature and sources of data to be transformed into information. There is an increase in the incoming new and even complex information sets into the systems environment of the firm.

4.3 Transactional IT

This section covers the actual to and fro of transactions/actions between BU's and how IT interacts with those transactions and puts charges on them. It also involves the maintenance of connectivity with regard to network connectivity, internet connectivity and connectivity of shared devices.

A major portion of this section can/may be dedicated to solving problems arising from lack of knowledge with regard to devices and software and wrong usage. One can refer to that process as trouble shooting.

So overall this side of IT can handle the likes of

- Operational Cost Management
- Trouble Shooting
- Connectivity Maintenance
- The Charging and Cost of IT.

4.4 IT Assets and Infrastructure

From an IT point of view, all devices, services and programs used to assist or enhance the day to day running and revenue generation of a firm, are considered IT ASSETS and make up parts of the overall IT architecture.

But this does not restrict itself to only what is already present or what has already been spent on. Neither is it about the maintenance of existing assets only. This category can render itself to suit the needs and aspirations of a firm. Thus it can include (but not only pertaining to)

- Costs: Incremental, Capacity and Sunk
- R & M of Existing Infrastructure
- New Initiatives
- Management of knowledge Content
- Purchases/Sales/Allocation
- Data Security and Integrity.

5 Conclusion

Investment in IT and the resulting communication and infrastructure assets needed requires an understanding of the strategic needs of the organization as well as how IT savvy the management teams are to extract value from data and the resulting big data. Accounting techniques such as DCF can ignore the likely costs of NOT investing in new technology. The choice of charging or budgeting for the recurrent IT expenditure, such as actual costs, does not provide incentives to the IT support to meet the needs of the intermediate customer (the SBU manager) nor be efficient in cost control. The Lord and Master charging system can be suitable for managers who are not IT savvy but the needs of present and future IT savvy customers may not be met. Outsourcing IT support, in the era of cloud technology does reduce the impact of large initial outlays while increasing reliability and stability of storage capacity. However, outsource agents are not the best agents for learning or change. The organization is reliant on the outsource agents and not masters of their own destiny. The proposed portfolio approach to IT investment and management requires a high level of IT savvy managers to leverage value from the IT assets and infrastructure. The existing investment by production companies in core activities is seventy percent; twenty percent for new business opportunities while ten percent invested in breakthrough or new market development [11]. Technology companies will invest more in transformational and less in core [11] to enable new or innovative products and markets to emerge.

The portfolio approach to IT investment does not fit the traditional norms of meeting timelines and budgeted cost. Strategies are dynamic and the IT investment and infrastructure must not only align to the strategy but also test the strategy. As

one CEO notes "any recommendation will be challenged, the reasoning challenged and the data challenged" [12]. Therefore, a continuous investment in IT assets and infrastructure to leverage the skills of "big data" scientists and IT savvy managers has become the competitive frontier.

References

1. Weill P, Aral S (2006) Generating premium returns on your IT investments. MIT Sloan Manage Rev 47:39–48
2. Gold RS (2004) Follow the money: IT finance and strategic alignment. Balanced Scorecard Rep March–April, 1–5
3. Marchand D, Peppard J (2013) Why IT fumbles analytics. Harvard Bus Rev Jan–Feb, 1–9
4. Yoo CS (2011) Cloud computing: architectural and policy implications. Rev Ind Organ 38:405–421
5. Weill P, Ross JW (2009) IT savvy: what top executives must know to go from pain to gain. Harvard Business School Publishing Corporation, Boston
6. Austin RD, Nolan RL, O'Donnell S (2009) The adventures of an it leader. Harvard Business School Press, Boston
7. Control Objectives for Information and related Technology (COBIT) provides detailed frameworks on IT governance. See www.isaca.org for detailed insights
8. Weill P, Ross JW (2005) A matrix approach to designing IT governance investments. MIT Sloan Manage Rev 48:26–34
9. Christensen C, Kaufman S, Shih W (2008) Innovation killers: how financial tools destroy your capacity to do new things. Harvard Bus Rev January, 98–105
10. Simons R (1995) Strategic levers of control. Harvard Business School Publishing Corporation, Boston
11. Nagji B, Tuff G (2011) Managing your innovation portfolio. Harvard Bus Rev May, 67–74
12. McGrath RG (2013) The end of competitive advantage: how to keep your strategy moving as fast as your business. Harvard Business Review Press, Boston, p 23
13 Kalpana RS, Atkinson A (1998) Advanced management accounting. Prentice Hall
14 Christensen CM, Kaufman S, Shih W (2008) Innovation killers: How financial tools destroy your capacity to do new things. Harvard Bus Rev January, 98–105

Virtual Test-Based Reliability Evaluation of Airborne Electronic Product

Cheng Qi, Li Chuanri and Guo Ying

Abstract In the development of airborne product reliability, the reliability of electronic products is an important factor to determine the reliability of aviation equipment. Based on the character and practicability of components in airborne product, virtual test technology is used to simulate the vibration and thermal stress that the airborne products are actually exposed to, and electronic component's life can be obtained through the reliability prediction software CALCE PWA. In terms of the failure life, the calculating methods of failure distribution for component level, board level and device level are proposed, which use single point distribution fitting, Monte Carlo random sampling and multipoint distribution fusion, to get the reliability indices such as MTBF etc. Based on these study, the evaluation methods relating to airborne product from component level to device level are established.

Keywords Virtual test · Airborne electronic product · Failure distribution · Reliability evaluation

1 Introduction

With the development of electronic technology, reliability and quality of electronic products are becoming more and more important across different fields. In civilian products field, reliability not only provides improvements for companies' product quality, but also enhances consumers' confidence. In the aerospace field, higher reliability is required due to the severe loss resulted from reliability accidents. Relevant statistics disclosed that the global plane crash occurred for more than 200 times in recent years, and consequently in the aerospace field, product reliability

C. Qi (✉) · L. Chuanri · G. Ying
School of Reliability and Systems Engineering, Beihang University, Beijing, China
e-mail: chengqiboy@163.com

L. Chuanri
e-mail: lichuanri@buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_112

cannot be ignored [1]. For the reliability assessment of aviation electronics, a large number of flight tests are not applicable to assess the reliability due to the cost and some other reasons. In the absence of experimental data, the full use of virtual simulation tests is fundamental for the assessment of reliability. Since relevant studies on reliability simulation have just started in China, reliability prediction and assessment of avionics product mainly depend on foreign research and software.

According to the development needs of domestic aviation electronics, this chapter use reliable virtual test as the basis to carry on simulation analysis and obtain the reliability parameters of electronic components through reliability simulation and prediction software, and then calculating the MTBF of board-level and device-level to complete the assessment of the reliability of aviation electronics according to failure distribution fitting of single point, distributed fusion of multipoint and so on. The issue start from the failure mechanism of components to fit the distribution and analyze the reliability indices step by step, from failure mechanism to component level, board level and device level, and finally achieved good results.

2 Reliability Virtual Test Method and Process

Airborne products mainly consist of the chassis, circuit boards, connectors and other components. The main part of failure is the component above the circuit board. In the avionics life cycle, it is known that the environmental conditions of that airborne electronic products withstanding are vibration and temperature changes based on actual test data. Therefore, thermal stress and vibration stress are the main load stresses that lead to the wear out failure of avionics. Reliability virtual test is based on physical failure theory and computer technology to establish digital model and analyse the effect of the applied load (vibration load and thermal load), thereby performing stress analysis and damage analysis [2]. In this chapter, virtual tests are used to analyse the vibration stress and temperature stress experienced by avionics in all-life cycle of product. The parameters calculated by virtual tests, such as load stress and ambient temperature, will be entered into the reliability prediction software, CALCE PWA, thus producing the estimated failure life data of components. A summary of our tests is available as Fig. 1 here.

2.1 Modelling and Simulation of Virtual Test

Electronic product design information is the key to the reliability virtual test. The more detailed on design information, the more reliable of the modelling. Meanwhile, modelling can be modified according to the type of test. For instance, during thermal simulation, some non-power devices, such as connectors and locking strips, can be deleted and some device of small size and quality can be negligible in the vibration simulation.

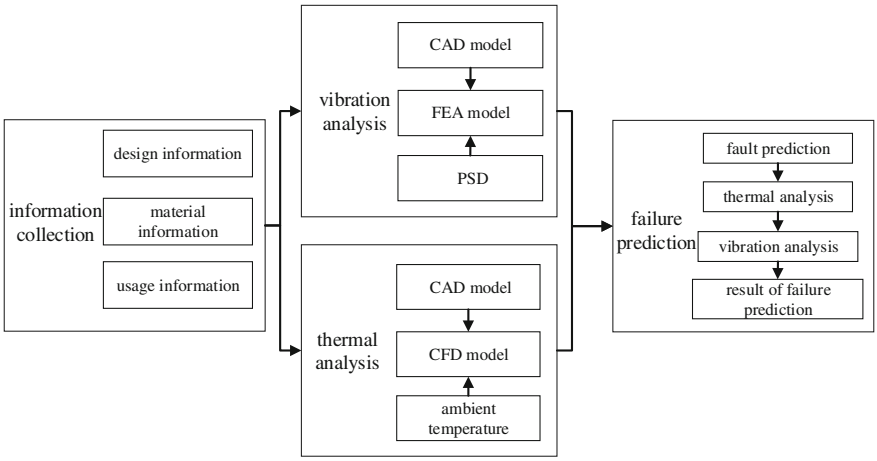


Fig. 1 Virtual test process

1. Thermal simulation modelling and analysis
   In this chapter, the software FLOTHERM is used for thermal analysis. FLOTHERM is developed by mature technology and numerical simulation of numerical heat transfer. Thermal analysis can be achieved from component level, board level and module level, device level. The model of thermal analysis can be obtained from SOLIDWORKS or designer and the material of thermal properties can be obtained through design document. After establishing CFD Digital Prototyping through FLOTHERM, the temperature distribution of the PCB shell temperature and components can be obtained by analysing the CFD Digital Prototyping in the FLOTHERM.
2. Vibration Simulation Modelling and Analysis
   Vibration simulation analysis mainly uses ANSYS. ANSYS is one of the largest general-purpose finite element analysis software that has integrated structure, fluid, electromagnetic, acoustic field and coupling field analysis. Vibration simulation modelling is generally constructed by digital prototyping, like CATIA and SOLIDWORKS models that are provided by design side. The size and structure of the product and related material properties can be obtained from CAD model and Design and Parts Manual accordingly. After establishing FEA Digital Prototyping in ANSYS system, ANSYS simulation can get a response PSD, stress and strain cloud, cloud displacement, acceleration cloud and so on.

2.2 Reliability Prediction

Theoretically, product reliability cannot be obtained after large numbers of life tests. However, in industrial production it is not a very economical way, as well as too late, to measure the reliability after products are produced. Therefore, it is very

necessary to control the reliability before manufacturing, which means that the reliability must be predicted in the product design stage, especially for large and complex systems. Software CALCE PWA, which was developed by University of Maryland, is based on physics of failure method of printed circuit boards. It provides an integrated design environment for the design of complex multi-layer printed circuit boards. First PWA determines the temperature and vibration stress at the joint, shell and bottom, and then calculates the failure rate of components and PCB [3], so as to prepare for the next step of reliability assessment. The steps of predicting failure are as follows:

(a) PCB modelling: First determine board size, number of layers and material, then determine the size, package, position, weight, power consumption, pins and other information of each component by querying manuals and design drawings. Ensure that the model is the same as the actual board
(b) Thermal stress analysis: Input the environmental temperature that obtained from simulation of FLOTHERM, then analysis the thermal stress of PCB combined with power consumption of each components.
(c) Vibration stress analysis: Input the PSD that obtained from simulation of ANSYS and then analyse the material information such as the position and weight of all components to get the PCB vibration stress.
(d) Simulation and prediction: First integrate the thermal stress and vibration stress, as well as the life profile that actually used, and set the Monte Carlo simulation times. Then CALCEPWA use materials, structures, process and performance parameters of the stress to create a digital model and analyse its failure mode, failure mechanism and effect to get its all potential failure points and corresponding physical model. Eventually it gets the large samples simulate failure time of each component under some failure mechanism through using stress damage to analyse every PCB.

3 Reliability Assessment

Reliability assessment is a kind of statistical inference of reliability characteristics which is based on the reliability of the product structure, product reliability life model and all relevant information. This chapter evaluates the reliability through the failure time matrix of components that CALCE PWA generated. The reliability level division of airborne electronic products is usually from component level to board level until device level. According to the result of simulate prediction, the failure is mainly caused by vibration and thermal failure. The chapter uses distributed fusion with the two failure mechanisms, fit the failure distribution and get the failure distribution of components. Then it use the failure matrix which can be obtained from the sampling method of Monte Carlo simulation to analyse reliability step by step and get the reliability parameters of device finally. The evaluation process is as Fig. 2.

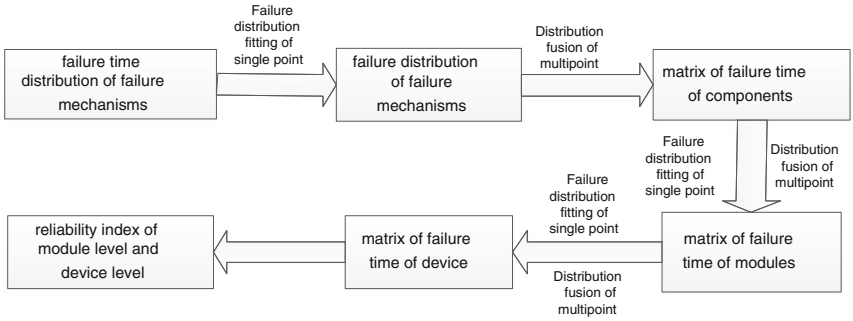


Fig. 2 Virtual test process

3.1 Failure Distribution Fitting of Single Point

According to the failure time matrix of one failure point that Monte Carlo simulation generated, distribution fitting is a method that using statistical method to fit the failure time distribution of potential failure point. Failure distribution fitting of single point can transform failure time to failure distribution, which can be prepared for the random sampling of Monte Carlo subsequently. It is very useful to use goodness-of-fit test to obtain the optimal failure distribution. Common distributions are normal distribution, exponential distribution, lognormal distribution and Weibull distribution. In engineering, parameters of Weibull distribution can express positive bias, negative bias, symmetry, and its capability of fitting curve is very excellent. Its basic function of reliability has a sealed analytical expression and it is very convenient for mathematical treatments. After double logarithmic transformation, it can be linearized so that the computer graphics processing and linear regression techniques can be easily and widely applied. Besides normal and exponential distribution can be seen as a special case of the Weibull distribution. Therefore it is very suitable to use Weibull distribution to fit the life distribution of board and device level in engineering [4].

3.2 Estimation Methods for 3-Parameter Weibull Distribution

In the past, common parameter estimation methods for 3-parameter Weibull distribution are graphing method, regression analysis method, third moment method and maximum likelihood estimator method. These methods are all exposed to the weaknesses of huge computation and low accuracy. This chapter use correlation coefficient method [5] to fit the life distribution, which is very simple, practical and suitable for the project.

Accumulative failure rate model of Weibull distribution is as follows

$$F(x) = 1 - \exp[-(\frac{x-\mu}{\sigma})^m] \tag{1}$$

After twice logarithmic transformation:

$$\ln(\ln(\frac{1}{1-F(x)})) = m\ \ln(x-\mu) - m\ \ln(\sigma) \tag{2}$$

Set

$$\left\{ \begin{array}{l} Y = \ln(\ln(\frac{1}{1-F(x)})) \\ X = \ln(x-\mu) \\ b = -m\ \ln(\sigma) \end{array} \right\} \tag{3}$$

Eq. (2) turn into

$$Y = mX + b \tag{4}$$

Equation (4) means that X and Y has a linear relationship. According to Eq. (2), when the estimated value of μ is correct, X and Y have the maximum correlation coefficient. So when the correlation coefficient R reaches the maximum value, the estimated value $\hat{\mu}$ of μ is the best estimated value. The specific procedures are as Fig. 3:

Given step t, correlation coefficient R vector can be obtained according to the procedure above. Taking R the largest number and the column number for V, so the estimated value of μ is equal to $V \times$ t. it is positional parameter. After positional parameter is obtained, we can use maximum likelihood estimation to get the value of m and σ.

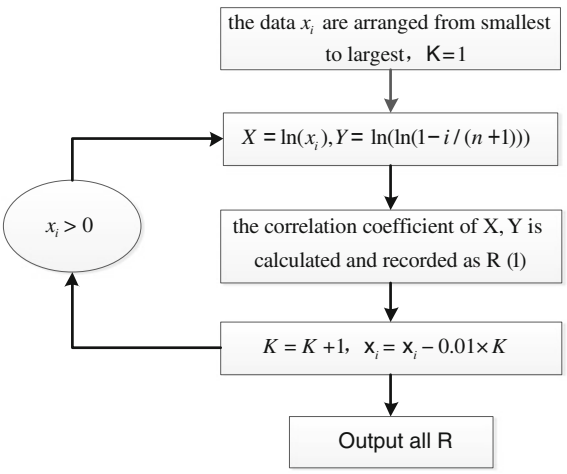


Fig. 3 The process of correlation coefficient method

In engineering, it is not accurate to express MTBF with mathematical mean. So mean of Weibull distribution is often used to replace MTBF [6]. The mean life is calculated as:

$$\text{MTBF} = \text{E(t)} = \mu + \sigma\Gamma\left(1 + \frac{1}{m}\right) \tag{5}$$

3.3 Distribution Fusion of Multipoint

Distribution fusion of multipoint fuses different failure points into general failure. Specifically, after fitting each failure element with failure distribution fitting of single point, failure distribution functions of each failure element is obtained. Assuming that all failure elements are independent, and failure distribution follow Weibull distribution. Using Monte Carlo simulation method for different failure distribution of each element to conduct a number of random sampling of failure time. After each sampling, the failure time is obtained based on competing failure, which is the failure time of module under a sampling. After simulating for N times, we can get a large number of failure time data of module [7].

4 Case Analysis

This case takes an airborne servo amplifier for example to introduce the procedure of reliability simulation and assessment. By thermal simulation, vibration simulation and failure prediction, a large number of matrix of failure time can be obtained. Through assessing the matrix of failure time, it can work out the MTBF of board and device eventually.

The device contains 6 PCB. The components of PCB are more than 2,700 including transistors, diodes, integrated circuits, chip capacitors, heat sink, and resistors and so on.

4.1 Thermal Analysis and Vibration Analysis

After the CAD model is simplified, it is inputted into FLOTHERM and ANSYS, and then the digital prototypes of CFD and FEA can be obtained after processing. The CAD model, digital prototypes of CFD and FEA are as Fig. 4.

Combined with the digital prototype of CFD, FLOTHERM set the power consumption, material, plate spacing, ambient temperature profile, etc. In ANSYS it set Young's modulus, density and Poisson's ratio combined with the digital prototype of FEA. Thus the results of stress analysis are obtained after thermal analysis and random vibration analysis. The results of thermal and vibration analysis are as Fig. 5.

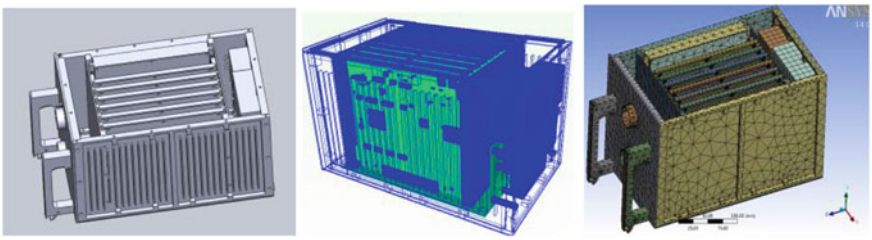


Fig. 4 The CAD model and digital prototypes of CFD and FEA

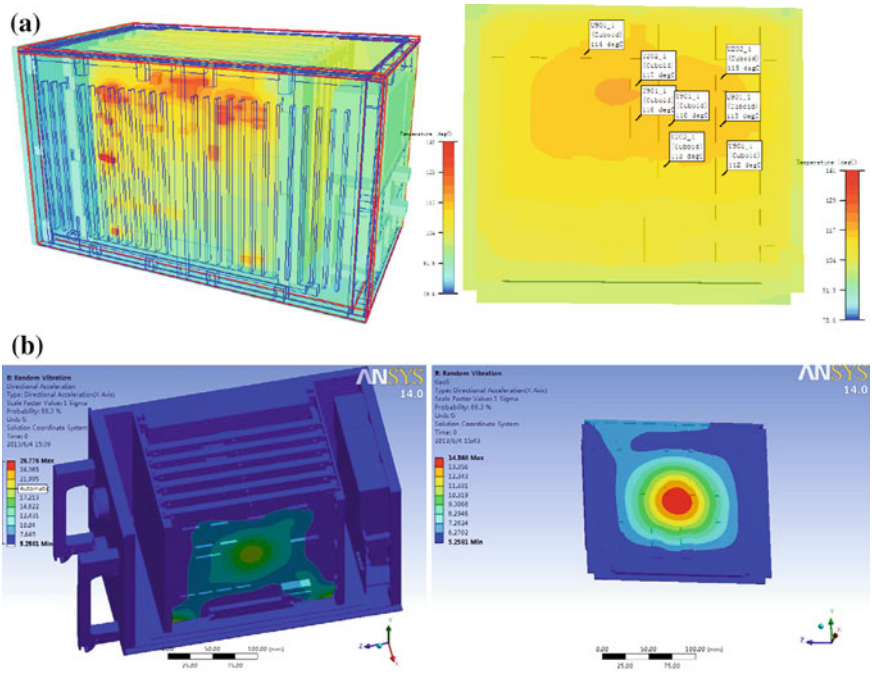


Fig. 5 **a** The thermal results of device and PCB. **b** The vibration results of device and PCB

4.2 Failure Prediction

Based on the collected information, the CALCE PWA establishes model for each PCB including pin, package, size, power consumption and position of every component in it. Then it need to determine the layers and copper content. One PCB model is as Fig. 6.

Input the PSD, which is obtained from vibration simulation and temperature distribution, which is obtained from thermal analysis into CALCE PWA, and calculate the effect of vibration and temperature. Combined with the effect of two

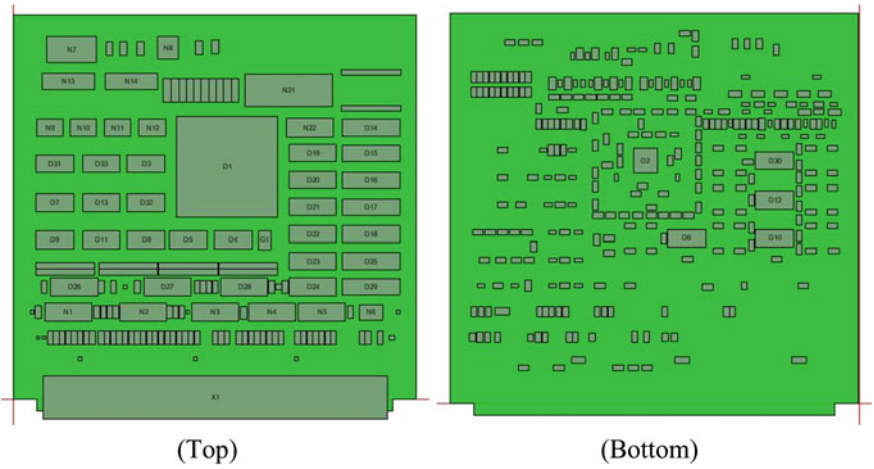


Fig. 6 The model of failure prediction

kind of stress, the times of Monte Carlo simulation are set up. And finally the failure mode (solder joints fracture, wire open, thermal fatigue, etc.), failure mechanism and a large number of failure time are obtained. The result of failure time is as Fig. 7.

4.3 Reliability Assessment

By reliability prediction, it can get a large number of matrixes of failure time under different failure mechanism. Using failure distribution fitting of single point to

Fig. 7 The result of failure prediction

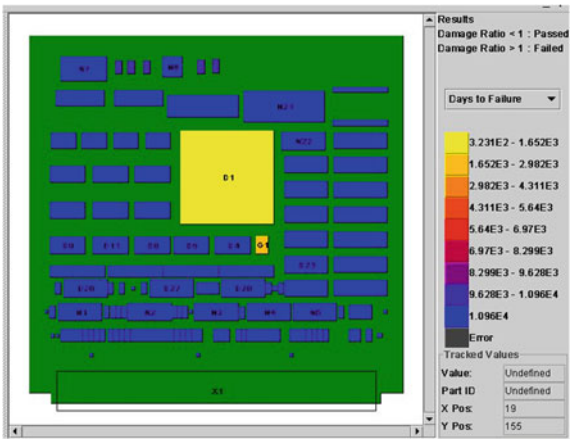


Table 1 The assessment result of each module and device

PCB id	Shape parameter	Scale parameter	Positional parameter	MTBF (h)
1	1.5924	1758	838.3	2415.1
2	3.925	842.06	1980.2	2742.6
3	1.854	1180.72	2773.61	3822.2
4	2.998	1585.24	2512.66	3928.2
5	2.532	943.56	1469.78	2307.2
6	1.25	1468.2	2136.1	3503.6
Device	2.691	1364.3	703.81	1916.9

conduct matrix of failure time and fit different failure distribution of each failure mechanism. Then use distribution fusion of multipoint to conduct the failure distribution obtained above, the number of times of sampling N = 1000. Thus the matrix of failure time of each component is obtained. Repeat the steps above to get the failure time distribution of each module as well as the device. According to failure time distribution, it can capture the MTBF of modules and device. Specific results of assessment are as Table 1.

5 Conclusion

This chapter uses reliability virtual test approach to analyse reliability assessment method that starts with failure mechanism level and calculates the MTBF through practical example. Compared with the traditional assessment methods, it starts from the analysis of the failure mechanism and solves the problem related to the different distribution of different mechanisms that traditional methods always ignore, and make the results more reasonable. Through this research, it establishes a kind of idea and method of assessment with good results and meets the domestic demand for avionics product.

References

1. Guo S, Hu Q (2010) Monte Carlo-based reliability analysis of electronic products. Electron Packag 10(5):33–35
2. Li Y, Dan K, Zeng C, Guo K (2007) Research on reliability virtual test and assessment of electronic product. Chin Qual 7:006
3. Zhu Y (2005) Technology and application of reliability physics of failure. Eng Appl 2.2:28–33
4. Wei P, He Y, Shu W, Yu H (2010) Application of weibull-based synthetically analysis method to life extension of airborne products. Aviat Maintenance Eng 1030
5. Zhao B, Wu S (2007) Parameter estimation methods for 3-parameter weibull distribution. Heat Treat Met 32(z1)

6. Xuan J, Lliu B, Chen C, Lei Y (2009) MTBF calculation study for aero-engine complete life data based on the weibull distribution model. Gas Turbine Exp Res 22(4)
7. Wan B, Fu G, Zou H (2011) Research on data processing method in reliability simulation and prediction of avionics. Electron Prod Reliab Environ Test 29(001):5–9

Feature Signal Extraction Based on Ensemble Empirical Mode Decomposition for Multi-fault Bearings

W. Guo, K.S. Wang, D. Wang and P.W. Tse

Abstract Multi-fault diagnosis for bearings is a challenge task. It is difficult to identify all the features from measured vibration signals when there is more than one bearing fault, especially, when some fault feature at the early stage is relatively weak and easily immersed in noise and other signals. The ensemble empirical mode decomposition (EEMD) method inherits the advantage of the popular empirical mode decomposition (EMD) method and can adaptively decompose a multi-component signal into a number of different bands of signal components called intrinsic mode functions (IMFs). In this chapter, the strategies of parameter optimization and signal component combination are combined with the normal EEMD to enhance its performance on signal processing. A vibration signal collected from a multi-fault bearing was used to verify the effectiveness of the enhanced EEMD method. The results demonstrate that the proposed method can accurately extract the feature signal; meanwhile, it makes the physical meaning of each IMF clear.

1 Introduction

For the vast majority of rotating machinery, rolling bearings are one of the most widely used and most likely to fail components. The fault of machinery reduces the production rate and increases the costs of production and maintenance, so knowledge of what, where and how faults occur is very important [1]. Research on rotating machinery multi-fault diagnosis is important to improve rotating machinery performance and safety [2].

W. Guo (✉) · K.S. Wang
School of Mechanical Electronic and Industrial Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, China
e-mail: w.guo.china@gmail.com

D. Wang · P.W. Tse
Department of Systems Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_113

Growing efforts are made to explore innovative methods to increase the performance of signal process for bearing fault diagnosis, such as, PCA [3], support vector machine classifier [4], blind source separation [5], wavelet analysis and hidden Markov model [6], and so on. However, at the early stage of bearing defects, the weak fault features are difficult to be identified. The weak bearing fault signal is easily to be overwhelmed by background noise and other vibration signals generated by strong defects and other machine components. Advanced signal processing methods are thus necessary to be developed for the identification of weak and multiple fault features.

Empirical mode decomposition (EMD) method [7, 8] is a popular adaptive signal processing method to decompose a nonlinear and non-stationary signal as the sum of some simpler signal components, termed as intrinsic mode functions (IMFs). This method is able to visualize signal energy spread between available frequencies locally in time, thus resembling the wavelet transform [9]. The EMD method has been proven to be effective for the fault diagnosis and structure health monitoring of rotating machinery. The mode mixing is one limitation of the EMD and degrades the accuracy of bearing fault diagnosis. Various methods for removing the mixed modes in the EMD method are published, in which Wu and Huang [10] introduced the white noise as the assistance for signal decomposition and presented a noise-assisted signal processing method, named ensemble EMD (EEMD) method. Its main idea is to use the property of the white noise to perturb the signal and enable the EMD method to visit all possible solutions in the finite neighborhood of the true final IMF [10]. Based on the property of zero mean for white noise, the added white noise can be cancelled each other out in the final ensemble mean if there are sufficient trials. Only the signal itself can survive in the final decomposition result. As of today, there are no general equations reported in the literature to guide the choice of EEMD parameters [11–14], especially the amplitude of the added white noise. In most of applications and modification, EEMD parameters were set to empirical values proposed by Wu and Huang in [10] or authors' empirical value. Žvokelj et al. [15, 16], Chang et al. [17], Zhang et al. [18], and Yeh et al. [19] independently introduced signal-to-noise ratio to select the noise amplitude. However, the assumption is the meaning signal or its power is known a priori. Niazy et al. [20] used an index, called relative root-mean-square error (RMSE), to evaluate the EEMD performances when setting different noise amplitudes. Guo and Tse [21] proposed a parameter optimization method for EEMD, in which relative RMSE was used to select the appropriate noise amplitude.

Fault diagnosis of multi-fault bearings is a challenging task. Multiple faults usually happen in one bearing simultaneously and have different characteristic frequency components. Due to the installation of measurement sensors and the property of faults, for accurate multi-fault diagnosis, it is difficult to identify the number, type, and level of faults, especially the relatively weak fault feature. The EEMD method, as an adaptive signal processing method, combining with parameter optimization, can decompose a raw vibration signal into a series of IMF. Another problem for the EEMD method is the dispersive signal components in the decomposition results, i.e. one signal component or a residue spreads in more than

one IMF, not a natural IMF. This is because the repeating decomposition process in EEMD. In this chapter, the optimal EEMD method would be combined with the strategy of signal component combination based on the cyclic coherence [22] to enhance the property of EEMD. This enhanced EEMD method is designed for the vibration signal analysis. An experiment of a multi-fault bearing and the corresponding fault diagnosis are conducted for investigating the performance of the enhanced EEMD method.

The remainder of this chapter is organized as follows. Section 2 briefly introduces the EEMD method. Section 3 applies the EEMD method with optimal parameters and signal combination to extract the bearing feature signal from the raw vibration signal collected from a multi-fault bearing. Experiment and fault diagnosis are conducted to verify the performance of the above signal processing method. Finally, conclusions are drawn in Sect. 4.

2 Ensemble Empirical Mode Decomposition

Ensemble Empirical Mode Decomposition (EEMD) method [10] is developed from the popular EMD method. It has been proven that EEMD is the improvement over the normal EMD method and solves one problem of mode mixing in EMD when the analyzed signal contains high-frequency intermittent oscillations. The procedure of the EEMD method is listed as follows:

Step 1 Initialize parameters: the ensemble number, N_E, and the amplitude of the added white noise, a, which is a fraction of the standard deviation of the signal to be analyzed. The index of the ensemble starts from 1, $m = 1$

Step 2 Generate the white-noise-added signal, $x_m = x + n_m$, where n_m is the added white noise with the pre-setting amplitude, and x is the signal to be analyzed

Step 3 Perform the m-th signal decomposition using the EMD method. The decomposition result is some IMFs, $c_{i,m}$, ($i = 1, 2, \ldots, N_{\mathrm{IMF}}$), and a non-zero low-order residue, r_m, where N_{IMF} is the number of IMFs obtained in each decomposition

Step 4 Repeat Steps 2 and 3 with $m = m + 1$ until the index m reaches the pre-setting ensemble number, N_E

Step 5 Calculate the ensemble means. Each final IMF, $\bar{c}_i$ (i = 1, 2, …, N_{IMF}) is the ensemble mean of the corresponding IMFs, $c_{i,m}$; and the final residue, $\bar{r}$, is the mean of all residues, r_m, where N_{IMF} is the number of IMFs obtained from each decomposition process.

Although each individual signal decomposition will generate a relatively noisy result, the added white noise is necessary to force the sifting process to visit all possible solutions in the finite neighborhood of real extreme and then generates different solutions for the final IMF [10]. Meanwhile, the zero mean of white noise is helpful for the cancellation of the added white noise in the final ensemble mean if

there are sufficient trails. Hence, only the signal itself can survive in the final decomposition result. This method not only solves the mode mixing problem, but also enhances the robustness of the decomposition results. It is applicable for analyzing and identifying the vibrations in rotating machines. For more details about EEMD method, please refer to Ref. [10].

3 Feature Signal Extraction for a Multi-fault Bearing

3.1 Parameter Optimization of EEMD

As mentioned in the previous sections, two critical parameters, the noise amplitude, A_N, (a fraction or noise level, L_N, of the standard deviation of the analyzed signal) and the ensemble number, N_E, need to be prescribed before signal decomposition if using the EEMD method. It would be ideal to automatically select the appropriate EEMD parameters, especially selecting the noise level, for the signal to be analyzed. An optimization method for EEMD parameters has been proposed in Ref. [21], in which an index, termed relative RMSE, was introduced to evaluate the performances of EEMD for various noise levels and hence to determine the optimal noise level for the analyzed signal; the index, signal-to-noise ratio, was used to evaluate the remaining noise in the decomposition result and then determine the appropriate ensemble number.

Assume that the original vibration signal, $x_o(k)$, is composed of the main component(s) related to the feature signal, the noise and some signal components with small correlation with the feature signal. In the decomposition results, the IMF, $c_{\max}(k)$, has the largest correlation coefficient with the original signal and contains the main signal component in the original signal. As a result, this IMF is selected as the feature signal for further signal analysis. The desired decomposition is to separate the feature signal from the noise and other irrelevant signal components. Relative RMSE is defined as the ratio between root-mean-square of the error and root-mean-square of the original signal, where the error is the difference between the original signal, $x_o(k)$, and the selected IMF, $c_{\max}(k)$. It is expressed as follows:

$$\text{RelativeRMSE} = \sqrt{\frac{\sum_{k=1}^{N} \left(x_o(k) - c\max(k)\right)^2}{\sum_{k=1}^{N} x_o^2(k)}}, \tag{1}$$

where N is the number of samples in the analyzed signal. If the relative RMSE is very small and close to zero, it indicates that the selected IMF, $c_{\max}(k)$, is close to the original signal; that is to say, the selected IMF, $c_{\max}(k)$, contains not only the main component of the original signal but also part of noise and other redundant signal components. Accordingly, the difference between the original signal and the

selected IMF is small and the desired decomposition has not been reached. However, there must be a value to maximize the relative RMSE. At this point, the selected IMF, $c_{\max}(k)$, contains only the main component and is separated from noise and other redundant components. Hence, the optimal noise level is the value that maximizes the relative RMSE. If the mean of the original signal is not zero, $x_o(k)$ in the denominator of Eq. (1) is replaced by the difference of the original signal and its mean value. It can remove the effect of non-zero mean and make the index independent of the mean value.

As for the ensemble number, it can be determined by using the signal-to-noise ratio after the optimal noise level is determined. The appropriate value of the ensemble number is the value that makes the remaining noise level minimal.

3.2 Combination of Related Signal Components

Due to the repeating decomposition operation in the EEMD method, there are two kinds of unsatisfied signal components in the decomposition results, one of which is some similar residues, the other of which is some successive IMFs with the same or similar frequency band. Such IMFs does not satisfy the definition of the IMF (a mono-component signal) and should exist in a single signal component. Therefore, a strategy is necessary to determine whether two successive IMFs are "related" and how to combine such IMFs. Considering that the signal components to be combined share similar features in the frequency domain, an index, cyclic coherence, was thus introduced to finish such a task. Cyclic coherence [22] between two signals, x and y (one is another shifted by a cyclic frequency), is to measure the linear spectral correlation at a specific frequency. The value of the cyclic coherence is in the range of 1 and 0. If the value of the cyclic coherence of two signals at a cyclic frequency is close to one, it indicates that these two signals are "strongly" jointly cyclo-stationary at that cyclic frequency. Otherwise, if the value of this index is close to zero, it indicates that these two signals to be analyzed share no cyclo-stationary at that cyclic frequency (even though they may be strongly coherent when the cyclic frequency is equal to zero). For the decomposition results obtained using the EEMD method, the cyclic coherence can be used to measure the spectral coherence of each two successive IMFs.

For the signal component to be analyzed, $c_i(t)$, its frequency spectrum is obtained by the fast Fourier transform and is denoted as $C_i(F)$ ($i = 1, 2, \ldots, N$). For two successive IMFs, $c_j(t)$ and $c_{j+1}(t)$, their spectral coherence, $\gamma_{j,j+1}(F)$ ($j = 1, \ldots, N-1$), is to measure the spectral linear dependence of their spectral components, $C_j(F)$ and $C_{j+1}(F)$, at the frequency of F. It is defined as the following equation:

$$\gamma_{j,j+1}(F) = \frac{C_j(F) \times C_{j+1}(F)}{\sqrt{\left(C_j(F) \times C_j(F)\right)\left(C_{j+1}(F) \times C_{j+1}(F)\right)}}. \tag{2}$$

Based on the definition of the cyclic coherence, the spectral coherence, $\bar{\gamma}_{j,j+1}$ ($j = 1, 2, \ldots, N-1$), is to measure the spectral coherence of two successive IMFs over whole frequency range. This index is calculated by using the following equation:

$$\bar{\gamma}_{j,j+1} = \frac{\sum_F C_j(F) \times C_{j+1}(F)}{\sqrt{\left(\sum_F C_j(F) \times C_j(F)\right)\left(\sum_F C_{j+1}(F) \times C_{j+1}(F)\right)}}. \tag{3}$$

The value of index $\bar{\gamma}_{j,j+1}$ is constrained to the range of zero and one. If the value of the index $\bar{\gamma}_{j,j+1}$ is close to 1, it indicates that these two successive IMFs are spectral coherence over the whole frequency range and can be combined into one natural IMF. If the value of the index $\bar{\gamma}_{j,j+1}$ is close to zero, the analyzed two signal components are spectral independent. If the value of the index is around 0.5, the spectral coherence of two signal components over the whole frequency can't be determined. Based on this calculation result of the index between each two IMFs, the decision of whether combining them can be made. When the value of the index is close to 1, the analyzed two IMFs belong to the one signal component and can be combined together. The corresponding procedure can be described as follows: firstly, calculate the spectral coherence $\bar{\gamma}_{j,j+1}$ of two IMFs, $c_j(t)$ and $c_{j+1}(t)$; if the value of the index is close to one, the signal component $c_j(t)$ is combined with the signal component $c_{j+1}(t)$ and replaces the original signal component $c_j(t)$, at the same time, update the sequence number of the IMFs and decrease the number of IMFs by one; otherwise, move to the next two successive IMFs; repeat this process until all of IMFs are considered.

3.3 Experiments and Analyses

In this sub-section, the enhanced EEMD method proposed above is applied to the multi-fault diagnosis of rotating machinery. A bearing with a multi-defect, in this case defects on the outer and the inner races, was considered in the experiment [21, 23]. Vibration signals were collected from the faulty bearing. The tested bearing (SKF 1206 EKTN9) was installed in a motor with a speed of 1,400 rpm. The sampling frequency for data acquisition was set to 80 kHz. The impulses generated by the inner race defect are rather weak and easily concealed by the impulses generated by the outer race defect and noise. It is thus difficult to identify such a weak bearing signal for accurate fault diagnosis. In addition, when the outer and the inner race defects exist in one bearing, the impacts individually generated by the faulty machine components influence each other; accordingly, the identified characteristic frequency may be different from the theoretical value obtained under the assumption that only one defect exists in the bearing.

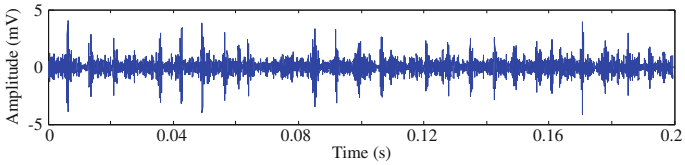


Fig. 1 A vibration signal measured from a multi-fault bearing, in which there are defects on the outer and the inner races

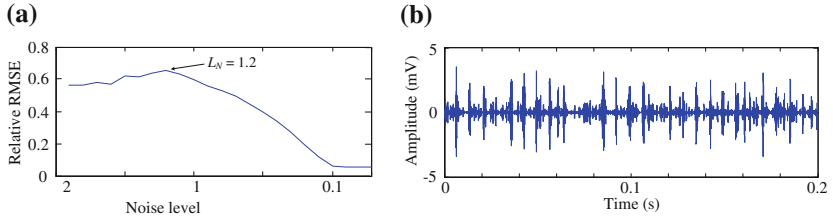


Fig. 2 Results using the parameter optimization: **a** Relative RMSEs when adding the white noise with various noise levels to the vibration signal from the bearing with a multiple defect (on the outer and inner races). The optimal noise level for this signal is 1.2. **b** The selected IMF (IMF2) when setting the noise level of 1.2 and its kurtosis value is 25.27

Figure 1 shows a vibration signal collected from the multi-fault bearing. After applying the parameter optimization method to this vibration signal, the relative RMSEs for different noise levels were obtained and are shown in Fig. 2a. The noise level corresponding to the maximum relative RMSE is 1.2 and is the optimal one. When setting this noise level, the selected IMF (IMF2) is shown in Fig. 2b. Its kurtosis value is 25.27. To compare with this decomposition result, other two non-optimal noise levels were set in the signal decomposition process and the results are shown in Fig. 3. Figure 3a, b show the selected IMFs when setting the noise levels, $L_N = 2$ and $L_N = 0.1$, respectively. The corresponding kurtosis values are 14.10 and 9.20, respectively.

Comparing the temporal waveform of IMFs in Fig. 2b with those in Fig. 3, it can be seen that some impulses are not involved in the selected IMFs. This is because impulses are partly distributed into other IMFs with smaller correlation coefficient with the original vibration signal. From the temporal waveforms, it can be seen that these extracted main signal components still contain much noise and the feature signals are not as obvious as that shown in Fig. 2b. This comparison demonstrates that the optimization method provides an appropriate noise level for the analyzed signal. After that, the appropriate ensemble number is set 140.

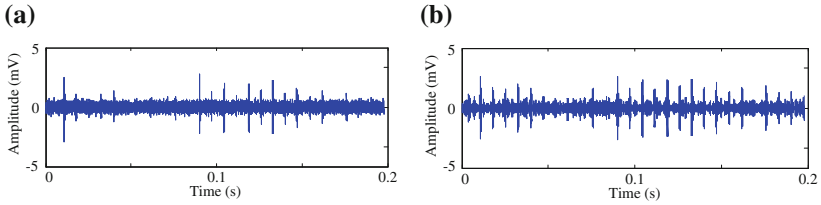


Fig. 3 The selected IMFs when setting non-optimal noise levels: **a** setting $L_N = 2$, and **b** setting $L_N = 0.1$. Their kurtosis values are 14.10 and 9.20, respectively

After computing the spectral coherence values of each two successive IMFs obtained by using the above optimal EEMD method, the results are shown in Table 1. As the table shows, the first two and the last three coherence values are close to ones. The former has relatively high value because some impulses are distributed into other IMFs, not in the main signal component. The latter are also caused by repeating signal decomposition and are residues in the decomposition process. Therefore, the first three IMFs and the last four IMFs are individually combined and generate the final signal decomposition results.

3.4 Validation—Fault Diagnosis

To validate the performance of the above enhanced EEMD method, the envelope spectral analysis is applied to the extracted bearing feature signal. For the tested bearing in the experiment, the theoretical ball pass frequency of the outer race (BPFO) and ball pass frequency of the inner race (BPFI) are 135 and 192 Hz, respectively. For the vibration signal collected from the multi-fault bearing (including the outer and the inner race defects), IMF1, IMF2 and IMF3 are combined as the bearing feature signal and used for fault diagnosis. The envelope spectrum of IMF2 is shown in Fig. 4 and reveals the identified BPFO (140 Hz) and its harmonics (2×, 3×, 4×, and 5× BPFO) along with the BPFI (209 Hz). In practice, there is always some sliding and slippage especially when the machine component in the bearing wears out. It leads to small frequency error between the theoretical and the identified one. Although the amplitude at BPFI is relatively small, this characteristic defect frequency (CDF) generated by the inner race defect was clearly observed in the frequency spectrum of the bearing signal. The envelope spectrum of the extracted bearing signal also demonstrates the effect of the interaction caused by multiple defective machine components in one bearing. The case of multi-fault diagnosis does not follow the assumption of CDF calculation and also contributes to the errors between the theoretical and the identified CDFs (BPFO and BPFI). However, it can be confirmed that the tested bearing has defects on two races.

Table 1 Spectral coherence values of each two successive IMFs obtained by using the normal EEMD method

	c_1, c_2	c_2, c_3	c_3, c_4	c_4, c_5	c_5, c_6	c_6, c_7	c_7, c_8	c_8, c_9	c_9, c_{10}	c_{11}, c_{12}	c_{12}, c_{13}
Spectral coherence	0.81	0.83	0.49	0.31	0.48	0.55	0.38	0.55	0.70	0.84	0.90

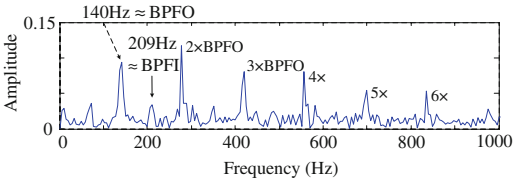


Fig. 4 Envelope spectrum of the extracted bearing signal

4 Conclusions

In this chapter, an enhanced EEMD method is proposed to extract bearing feature signal for fault diagnosis of multi-fault bearings, in which the parameter optimization method is used to automatically select the appropriate EEMD parameters for the vibration signal to be analyzed, and the index spectral coherence is used to determine whether combine successive IMFs with the similar frequency band. The former is to solve the problem of empirical parameter setting, and the latter is to remove the problem of disperse signal components in the decomposition results, which is caused by the repeating decomposition process in the EEMD method. Experimental results demonstrate that the enhanced EEMD method is effective for extracting the feature signal of the multi-fault bearing. Furthermore, the signal components are distributed in right IMFs, and their meaning is clearer than those obtained using the normal EEMD method. The envelope spectral analysis on the extracted feature signal reveals the characteristic defect frequencies hidden in the bearing signal and determines the fault types of the tested bearing.

Acknowledgment The work that is described in this chapter is fully supported by the Fundamental Research Funds for the Central Universities (Project No. ZYGX2012J105 and Project No. ZYGX2013J094). The authors wish to thank Dr. Peter W. Tse in City University of Hong Kong for the allowance of using experimental bearing data. We appreciate three anonymous reviewers for their valuable comments and suggestions for improving this chapter.

References

1. Hajiaghajanim M, Toliyat HA (2004) Advanced fault diagnosis of a DC motor. IEEE Trans Energy Convers 19(1):60–65
2. Jiang H, Li C, Li H (2013) An improved EEMD with multi-wavelet packet for rotating machinery multi-fault diagnosis. Mech Syst Signal Process 36(2):225–239
3. Li Z, Yan X, Yuan C, Peng Z, Li L (2011) Virtual prototype and experimental research on gear multi-fault diagnosis using wavelet-autoregressive model and principal component analysis method. Mech Syst Signal Process 25(7):2589–2607
4. Tang X, Zhuang L, Cai J, Li C (2010) Multi-fault classification based on support vector machine trained by chaos particle swarm optimization. Knowl-Based Syst 23(5):486–490
5. Jing J, Meng G (2009) A novel method for multi-fault diagnosis of rotor system. Mech Machine Theory 44(4):697–709

6. Purushotham V, Narayanan S, Prasad SAN (2005) Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition. NDT and E Int 38(8):654–664
7. Huang NE, Shen Z, Long R et al (1998) The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. Proc Royal Soc London Ser A-Math Phys Eng Sci 1998:903–995
8. Huang NE, Shen Z, Long SR (1999) A new view of nonlinear water waves—the Hilbert spectrum. Ann Rev Fluid Mech 31:417–457
9. Feldman M (2009) Analytical basics of the EMD: Two harmonics decomposition. Mech Syst Signal Process 23(7):2059–2071
10. Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advan Adapt Data Anal 1(1):1–41
11. Lei Y, Zuo MJ (2009) Fault diagnosis of rotating machinery using an improved HHT based on EEMD and sensitive IMFs. Measure Sci Technol 20(12):125701 (12 pp)
12. Lei Y, He Z, Zi Y (2011) EEMD method and WNN for fault diagnosis of locomotive roller bearings. Expert Syst Appl 38(6):7334–7341
13. Lei Y, He Z, Zi Y (2009) Application of the EEMD method to rotor fault diagnosis of rotating machinery. Mech Syst Signal Process 23(4):1327–1338
14. Zhou Y, Tao T, Mei X, Jiang G, Sun N (2011) Feed-axis gearbox condition monitoring using built-in position sensors and EEMD method. Robot Comput-Integr Manufact 27(4):785–793
15. Žvokelj M, Zupan S, Prebil I (2011) Non-linear multivariate and multiscale monitoring and signal denoising strategy using Kernel Principal Component Analysis combined with Ensemble Empirical Mode Decomposition method. Mech Syst Signal Process 25 (7):2631–2653
16. Žvokelj M, Zupan S, Prebil I (2010) Multivariate and multiscale monitoring of large-size low-speed bearings using Ensemble Empirical Mode Decomposition method combined with Principal Component Analysis. Mech Syst Signal Process 24(4):1049–1067
17. Chang KM, Liu SH (2011) Gaussian noise filtering from ECG by Wiener filter and ensemble empirical mode decomposition. J Signal Process 64(2):249–264
18. Zhang J, Yan R, Gao RX, Feng Z (2010) Performance enhancement of ensemble empirical mode decomposition. Mech Syst Signal Process 24(7):2104–2123
19. Yeh JR, Shieh JS, Huang NE (2010) Complementary ensemble empirical mode decomposition: a novel noise enhanced data analysis method. Advan Adapt Data Analy 2 (2):135–156
20. Niazy RK, Beckmann CF, Brady JM, Smith SM (2009) Performance evaluation of ensemble empirical mode decomposition. Advan Adapt Data Analy 1(2):231–242
21. Guo W, Tse PW (2013) A novel signal compression method based on optimal ensemble empirical mode decomposition for bearing vibration signals. J Sound Vibrat 332(2):423–441
22. Antoni J (2007) Cyclic spectral analysis in practice. Mech Syst Signal Process 21(2):597–630
23. Guo W, Tse PW, Djordjevich A (2012) Faulty bearing signal recovery from large noise using a hybrid method based on spectral kurtosis and ensemble empirical mode decomposition. Measurement 45(5):1308–1322

The Influence of Corrosion Test on Performances of Printed Circuit Board Coatings

Chengyu Ju, Xiaohui Wang, Run Zhu and Xiaoming Ren

Abstract In this chapter, the failure of acrylic coatings has been studied systematically by fungus test, salt fog test and humidity test, and then we designed a laboratory test to simulate marine climate. We analyze failure of acrylic coating and compare laboratory test with field test by SEM (scanning electron microscope) and EDS (energy dispersive spectrometer), The results show that there is a good correlation between laboratory test and field test in failure mechanism and failure mode.

Keywords Acrylic coating · Fungus test · Salt fog test · Humidity test · Insulation resistance

1 Introduction

Printed circuit board is the carrier of electronic components which is widely used in areas such as airborne electronic equipment. Its coating is the first barrier for protection circuit. Once the coating is damaged, the performances of printed circuit board will decrease rapidly until failure, therefore, the coating has important roles in the entire printed circuit board. When the coating works in high temperature, high humidity [1] or in marine environment [2], it will lead to degradation [3, 4] in mechanical [5], thermal, porosity [6] and electrical insulation performance. So we designed a laboratory test to simulate marine climate. We analyze failure of acrylic coating [7] and compare laboratory test with field test by SEM (scanning electron microscope) and EDS (energy dispersive spectrometer), The results show that there is a good correlation between laboratory test and field test in failure mechanism and failure mode.

C. Ju (✉) · X. Wang · R. Zhu · X. Ren
Reliability and Systems Engineering, Beihang University, Beijing 100191, China
e-mail: juchengyu@139.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_114

2 Test Methods and Test Scheme

2.1 Test Material

Acrylic coating on printed circuit board.

2.2 Test Methods

Clean the sample by anhydrous ethanol before the experiment and then dry it. Test methods are as follows in Fig. 1.

2.3 Test Conditions

2.3.1 Humidity Test

Y751C-12® alternating temperature humidity test chamber is used in the laboratory test. Test conditions is set according to GJB150.9 [8], test conditions are as follows: low temperature at 30 °C ± 20 °C, high temperature at 60 °C ± 2 °C and the relative humidity at 90–98 %, test period: 10 days.

2.3.2 Fungus Test

QDF-10KA® fungus test chamber is used. Test conditions are according to GJB150.10 [9]. The temperature is 30 °C ± 1 °C and the relative humidity is 95 ± 5 %, test period: 28 days.

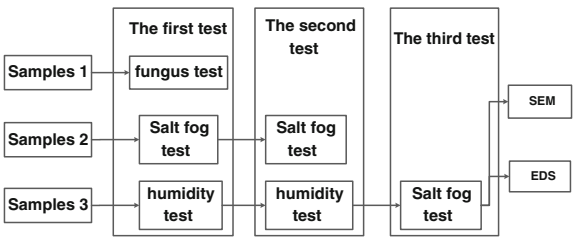


Fig. 1 The test methods of the coating

2.3.3 Salt Fog Test

NQ-1000A® salt spray cabinet is used. Test conditions are according to GJB150.11 [10]. Test conditions are as follows: NaCl concentration: 5 %, pH: 6.5–7.2. The temperature is 35 °C ± 2 °C and the fallout rate is 1.5 ml/80(cm^2·h), test period: 96-h.

2.3.4 Field Test in Wanning Station

The test station is located in tropical marine climate. The annual average temperature is 24.6 °C, the average humidity is 86 % and average annual rainfall is 1,942 mm. The average concentration of Cl^- is 9,729 mg/m^3. Test period is 1 year; the observed time is 1st month, 2nd months, 3rd months, 6th months, 9th months and 12th months. The measurement is insulation resistance of the coating according to GJB 360A-1996 [11]. The voltage of apparatus is 500 ± 10 % V DC. We analyze the sample by using SEM (scanning electron microscope) and EDS (energy dispersive spectrometer) after 12 months experiment.

2.4 Measurement of Insulation Resistance

According to the standard QJ519A-19995 [12], the GPI-735A® testing instrument is used to test insulation resistance of the coating. Its testing voltage is 500 V and the measuring ranges of the apparatus is 1–9,900 MΩ. If it is beyond quantum, it only can provide whether it is smaller than the minimum range or larger than the greatest range.

2.5 Measurement of Moisture Absorption

BS-124S® electronic weighing instrument is used. Measurement accuracy is 0.0001 g. After cleaning and drying the sample, weigh the sample by electronic scale and records G_0. After the test, we remove the water with filter paper and dry the coating. Then weigh the sample and records G_1. At last calculate the rate of moisture absorption ρ. The calculation formula is as follows:

$$\rho = (G_1 - G_0)/G_0 \tag{1}$$

3 Experimental Results and Analysis

3.1 The First Test Results

3.1.1 Fungus Test

The result rating the extent of microbial growth in the coating after fungus test according to GJB150.10 [9] shows substrate is devoid of microbial growth. Its rating is level 0. The insulation resistance value is greater than 9,900 MΩ and the rate of moisture absorption is 0.05042 %.

3.1.2 Salt Fog Test

Analysis of the coating after salt fog test according to GB/T 1766-2008 [13], acrylic coating becomes pulverization in appearance. The insulation resistance value is greater than 9,900 MΩ and the rate of moisture absorption is 0.5320 %.

3.1.3 Humidity Test

Analysis of the coating after humidity test according to GB/T 1766-2008 [13], there is a discoloration on the coating in appearance. The insulation resistance value is greater than 9,900 MΩ and the rate of moisture absorption is 1.0553 %.

3.1.4 Data Analysis

(1) There is a good anti-mildew property for acrylic coating according to the results of fungus test.
(2) The moisture absorption has been improved after the humidity. The main reason is as follows: as the temperature changes, condensation phenomenon will appear. It will prompt the coating to absorb moisture. What is more, the change of temperature can affect performance of the coating and lead to aging phenomenon. It can also promote moisture absorption of the coating.
(3) There is a little degeneracy phenomenon on the coating after salt fog test and humidity test by the appearance. But the insulation resistance is greater than 9,900 MΩ.
(4) According to the analysis above, there is no obvious influence on the insulation resistance of the coating after corrosion test. Therefore, the second salt fog test and humidity test are put on. The test condition is the same with the first test in order to strengthen the impact on the coating by corrosion test.

3.2 The Second Test Results

Fungus test is got rid in the second test. There are two reasons: First, Fungus test has a very small influence on the coating added fungicide and the conditions are too mild, so it has little impact on appearance, moisture absorption and insulation resistance. Second, Wanning station belongs to tropical marine climate. It is difficult for microbial growth in the higher salt content environment. Therefore, humidity and salt fog are two main factors.

3.2.1 Salt Fog Test → Salt Fog Test

The second salt fog test is put on. The condition of the test is the same as the first salt fog test. The result shows that insulation resistance value is still greater than 9,900 MΩ and the rate of moisture absorption is 1.1826 %.

3.2.2 Humidity Test → Humidity Test

The second humidity is put on. The conditions of the test are the same with the first humidity test. The result shows insulation resistance value of the coating is less than the minimum range (10 MΩ) in the edge of the solder paste and through-hole on printed circuit board. Obviously this part of the coating is failure. The rate of moisture absorption is 1.1005 %.

3.2.3 Data Analysis

(1) From the experimental phenomena we can conclude that humidity test have larger effect on the coating than salt fog test in the second test. It mainly because humidity test has a long time and there exist temperature cycling, therefore, it has a larger effect on the coating. It improves very slowly after the rate of moisture absorption exceeds 1 %. Therefore that moisture absorption of the coating has reached saturation.
(2) From the results of the experiment we can also conclude that the edge of the solder paste and through-hole on the coating are weaknesses in printed circuit board. Continuous data between 1 and 9,900 MΩ of insulation resistance is not detected. So the value of insulation resistance falls suddenly means the rapid failure of the coating.

3.3 The Third Test Results

3.3.1 Humidity Test → Humidity Test → Salt Spray Test

To simulate the real environmental better and evaluate the performance of acrylic coating by corrosion test, we do salt fog test again after the humidity test twice. The result shows insulation resistance value of the coating is less than the minimum range (10 MΩ) in the edge and surface of the solder paste and through-hole of the printed circuit board. The rate of moisture absorption is 1.2926 %.

Next we compare the coating by humidity test → humidity test → salt spray test with field test on corrosion morphology, corrosion composition, and insulation resistance etc.

(1) The results of the third test and field test are alike on corrosion morphology of the pin coating, all the coating are covered with large amount of corrosive substance in the pin of the component, both of their coatings generate cracks and the coatings on the corrosion position are loose and porous (Figs. 2 and 3), what is more, corrosion products of the third test and field test are alike by EDS, we find the coating on the edge of the pins contains many contents of Sn and Pb and a little bit of Cl. That prove the solder is corroded because the contents of Sn and Pb are composition of solder under the coating. It can infer the two elements first dissolve and diffuse and then they flew into the coating.

(2) All the coating emerge cracking and dropping after the third test and field test on the corners and through-hole (Figs. 4 and 5), because of the influence of edge effects between components and substrate of printed circuit board, when we spray painting on printed circuit board, it is hard to keep the coating

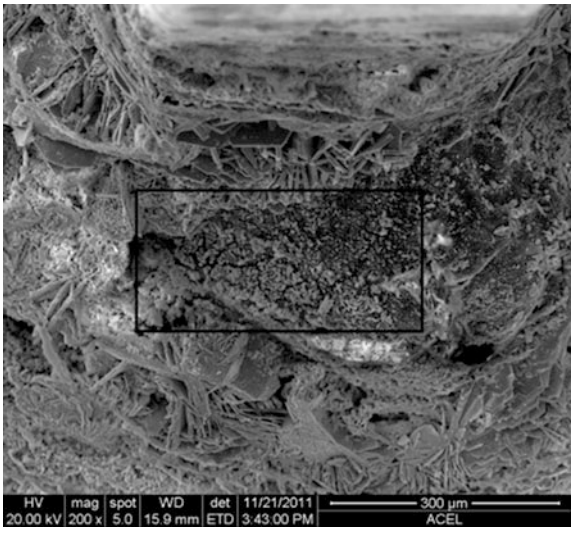


Fig. 2 SEM image on the edge of the pins after the third test

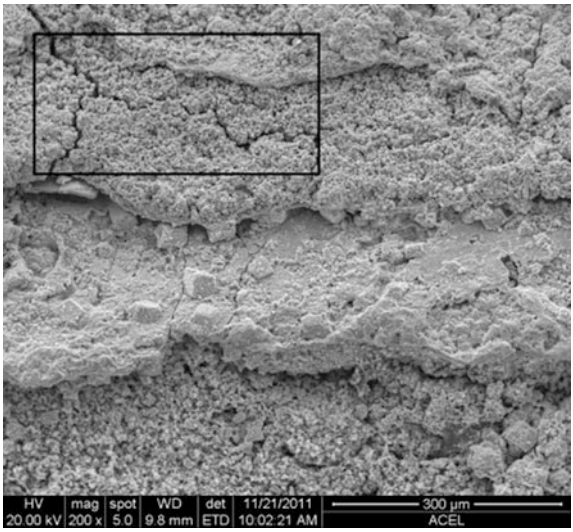


Fig. 3 SEM image on the edge of the pins for the 12th month field test

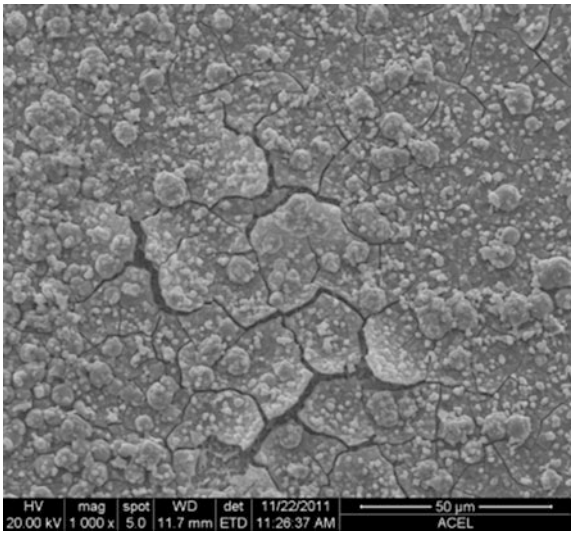


Fig. 4 SEM image on the coating after the third test

uniform. The thin parts will be destroyed at first. Humidity accumulated will generate the coating swelling, deformation and internal stress for a long time; serious cracking and dropping phenomenon of the coating is emerged.
Therefore there is a good correlation between laboratory test and field test in failure mechanism and failure mode.

(3) We obtained few accuracy data on insulating resistance for the measuring ranges of the apparatus by the laboratory test. But we can still infer that the insulating resistance value of the coating stays stable stage firstly after a series

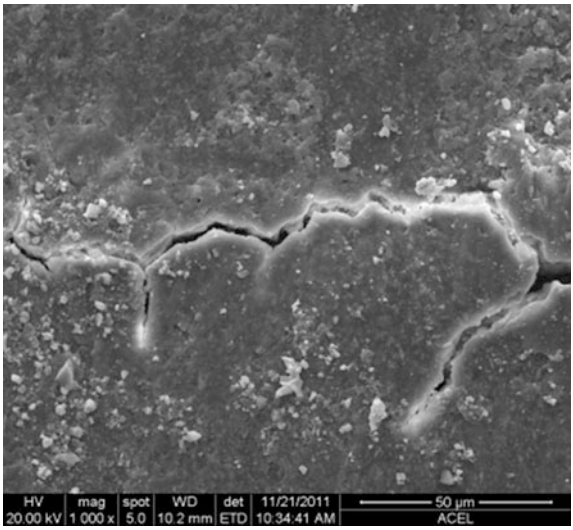


Fig. 5 SEM image on the coating for the 12th month after field test

of test, and then it falls suddenly due to the failure of the coating. Figure 6 shows the trend of insulating resistance value by the field test. The insulating resistance value almost did not change at all and the performance of the coating is good on the first 2 months. As time passes, the insulation resistance value keeps decreasing and become more and more quickly. After the ninth month, it decreased sharply. The phenomena are similar to those by the laboratory test.

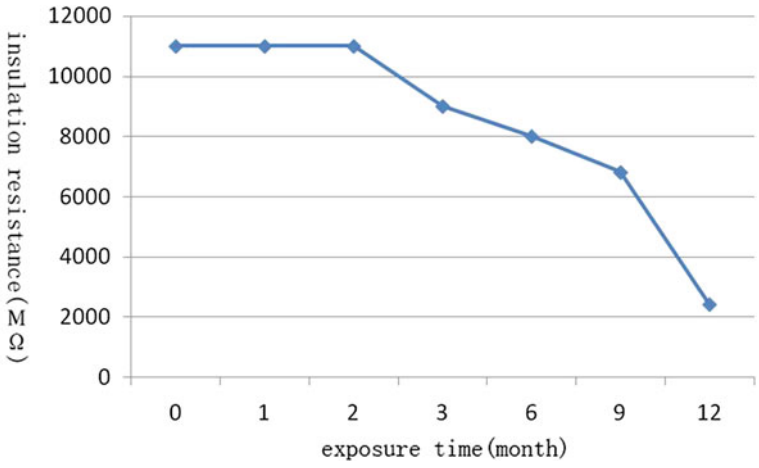


Fig. 6 Electrical capacity trend of the coating by the field test

4 Conclusion

(1) We discovered that humidity test play the most important role on the rate of moisture absorption, then salt fog test, the last fungus test during the corrosion tests.
(2) The combination test of humidity test → humidity test → salt fog test can simulate the field test well, they are relatively similar on corrosion morphology, corrosion product and so on.
(3) We find that the value of insulation resistance falls suddenly by the laboratory test and field test. If we can find out the time before it happen, we can predict its failure life. To find the failure life will be a research direction of this subject.
(4) This chapter provides many detection methods about the coating on Printed circuit board. The weakness such as the edge of the solder paste and through-hole of the printed circuit board are confirmed. What is more, there is a good correlation between laboratory test and field test in failure mechanism and failure mode.

References

1. Dehri I, Erbil M (2000) The effect of relative humidity on the atmospheric corrosion of defective organic coating materials: an EIS study with a new approach. Corros Sci 42 (6):969–978
2. Xing Li, Xiaohui Wang (2006) Overview of Three-Proof Design on Carrier-Based Aircraft. Equip Environ Eng 3(4):12–15
3. Xue TJ, Wilkie CA (1997) Thermal degradation of poly (styrene-g-acrylonitrile). Polym Degrad Stab 56(1):109–113
4. Gulmine JV, Janissek PR, Heise HM, Akcelrud L (2003) Degradation profile of polyethylene after artificial accelerated weathering. Polym Degrad Stab 79(3):385–397
5. Sangaj NS, Malshe VC (2004) Permeability of polymers in protective organic coatings. Prog Org Coat 50(1):28–39
6. Kosek JR, DuPont JN, Marder AR (1995) Effect of porosity on resistance of epoxy coatings to cold-wall blistering. Corrosion 51(11):861–871
7. Perrin FX, Irigoyen M, Aragon E, Vernet JL (2000) Artificial aging of acrylurethane and alkyd paints: a micro-ATR spectroscopic study. Polym Degrad Stab 70(3):469–475
8. GJB150.9 Environmental test methods for military equipments-Damp heat test. Commission on Science, Technology, and Industry for National Defense, China
9. GJB150.10 Environmental test methods for military equipments-Fungus test. Commission on Science, Technology, and Industry for National Defense, China
10. GJB150.11 Environmental test methods for military equipments-Salt fog test. Commission on Science, Technology, and Industry for National Defense, China
11. GJB 360A-1996 Test methods for electronic and electrical component parts. Commission on Science, Technology, and Industry for National Defense, China
12. QJ519A-1999 Test methods for printed circuit board. Aerospace Science and Technology Corporation, China
13. GB/T 1766-2008 Paints and varnishes-Rating schemes of degradation of coats. Standardization Administration of the People's Republic of China

Comparative Analysis of Printed Circuit Board Coating on Corrosion Test

Chengyu Ju, Xiaohui Wang, Run Zhu and Xiaoming Ren

Abstract In this paper, four kinds of coating (acrylic coating, polyurethane coating, silicone coating, ParyleneC coating) on printed circuit board have been studied in terms of moisture absorption, insulation resistance and surface morphology by fungus test, salt fog test and humidity test. The results show humidity test has largely effect on coating than salt fog test; the smallest effect is fungus test. ParyleneC coating is the best material by the three corrosion test, the second is silicone coating, the third is polyurethane coating, and the last one is acrylic coating. What is more, the weakest parts on acrylic coating are the edge of the solder paste, through-hole and so on.

Keywords Coating · Fungus test · Salt fog test · Humidity test · Insulation resistance · Moisture absorption · AFM (atomic force microscopy)

1 Instruction

Printed circuit board is the carrier of electronic components which is widely used in areas such as airborne electronic equipment. Its coating is the first barrier for protection circuit. Once the coating is damaged, the performances of printed circuit board will decrease rapidly until failure [1], therefore, the coating has important roles in the entire printed circuit board. Lots of work in terms of environmental behavior and failure mechanism of coating has been done and remarkable progress has been made in recent years [2–5]. Now AR (acrylic acid), SR (organ silicone), UR(Polyurethane), XY (Parylene) and ER(modified epoxy resin) are widely applied as protective coatings. There are large differences in these types of coatings on the selection of the material and its performances [6]. In this paper, we make comparison test by GJB, so that we can choose materials that is used on the surface of PCB better.

C. Ju (✉) · X. Wang · R. Zhu · X. Ren
Reliability and Systems Engineering, Beihang University, Beijing 100191, China
e-mail: juchengyu@139.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_115

2 Test Method

2.1 Testing Materials

Acrylic coating on printed circuit board, Polyurethane coating, Silicone coating, ParyleneC coating.

2.2 Test Scheme

Clean the sample by anhydrous ethanol before the experiment and then dry it. Test methods are as follows in Fig. 1.

2.3 Test Conditions

2.3.1 Humidity Test

Y751C-12® alternating temperature humidity test chamber is used in the laboratory test. Test conditions is set according to GJB150.9 [7], test conditions are as follows: low temperature at 30 °C ± 2 °C, high temperature at 60 °C ± 2 °C and the relative humidity at 90–98 %, test period: 10 days.

2.3.2 Fungus Test

QDF-10KA® fungus test chamber is used. Test conditions are according to GJB150.10 [8]. The temperature is 30 °C ± 1 °C and the relative humidity is 95 % ± 5 %, test period: 28 days.

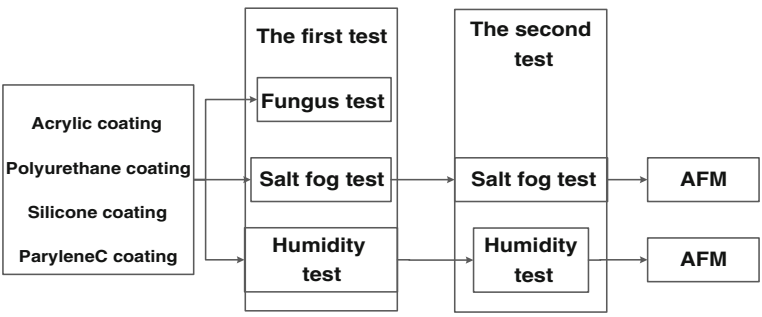


Fig. 1 Test Scheme of the Coatings

2.3.3 Salt Fog Test

NQ-1000A® salt spray cabinet is used. Test conditions are according to GJB150.11 [9]. Test conditions are as follows: NaCl concentration: 5 %, pH: 6.5–7.2, the temperature is 35 °C ± 2 °C and the fall out rate is 1.5 ml/80 ($cm^2 \cdot h$), test period: 96-h.

2.4 Sample Performance Test

2.4.1 Measurement of Insulation Resistance

According to the standard QJ519A-19995 [10], the GPI-735A® testing instrument is used to test insulation resistance of the coating. Its testing voltage is 500 V and the measuring ranges of the apparatus is 1–9,900 MΩ. If it is beyond quantum, it only can provide whether it is smaller than the minimum range or larger than the greatest range.

2.4.2 Measurement of Moisture Absorption

BS-124S® electronic weighing instrument is used. Measurement accuracy is 0.0001 g. After cleaning and drying the sample, weigh the sample by electronic scale and records G_0. After the test, we remove the water with filter paper and dry the coating. Then weigh the sample and records G_1. At last calculate the rate of moisture absorption ρ. The calculation formula is as follows:

$$\rho = (G_1 - G_0)/G_0$$

3 Experimental Results and Analysis

3.1 The First Test Results

3.1.1 Fungus Test

The result rating the extent of microbial growth in the coating after fungus test according to GJB150.10 [8] shows as follows.

From Table 1 we know that substrate is devoid of microbial growth. Its rating is level 0. The insulation resistance value is greater than 9,900 MΩ, which proves that the antifungal property and insulating property of these coating materials are fairly good. Moisture absorption rate from largest to smallest is that: Polyurethane coating > Silicone coating > Acrylic coating > ParleneC coating. And moisture absorption rate of the four coating is relatively small. The reason is that the added

Table 1 The results of fungus testing

Material type	Ratings	Moisture absorption (%)	Insulation resistance (MΩ)
Acrylic	0	0.05042	>9,900
Polyurethane	0	0.06352	>9,900
Silicone	0	0.05299	>9,900
ParyleneC	0	0.01053	>9,900

mildew protective coating makes it difficult for fungus to grow, so the damage to coating becomes minimal. And fungus test conditions are mild which make coating less affected by the temperature and humidity. So the moisture absorption rate is very low.

3.1.2 Salt Fog Test

Analysis of the coating after salt fog test according to GB/T 1766-2008 [11], acrylic coating becomes pulverization and polyurethane coating becomes discoloration in appearance, but the extent is not serious. Concrete results are shown in Table 2:

From Table 2, we found that after salt fog test, all four of insulation resistance of coatings for more than 9,900 MΩ, which proved their good insulations. Moisture absorption rate from largest to smallest is: Polyurethane coating > Silicone coating > Acrylic coating > ParleneC coating. From the salt fog test results, we noticed that compared to fungus test, the four kinds of coated moisture absorption rate in salt fog test improves nearly 10 times, but the sort of moisture absorption rate does not change. As can be seen, salt fog has more effect on moisture absorption rate of these printed circuit board coatings.

3.1.3 Humidity Test

Analysis of the coating after humidity test according to GB/T 1766-2008 [11], there is a discoloration on the coating in appearance. Concrete results are shown in Table 3:

Table 2 The results of salt fog test

Coating types	Moisture absorption rate (%)	Insulation resistance (MΩ)
Acrylic	0.5320	>9,900
Polyurethane	0.6523	>9,900
Silicone	0.5964	>9,900
ParyleneC	0.1055	>9,900

Table 3 The results of humidity test

Coating types	Moisture absorption rate (%)	Insulation resistance (MΩ)
Acrylic	1.0553	>9,900
Polyurethane	1.2424	>9,900
Silicone	0.8067	>9,900
ParyleneC	0.1089	>9,900

From the experimental phenomena, we find that, all four kinds of insulation resistance of coatings are more than 9,900 MΩ after humidity test, which proved its good insulation. Moisture absorption rate from largest to smallest is: Polyurethane coating > Acrylic coating > Silicone coating > ParleneC coating. Compared to salt fog test, Acrylic and polyurethane coatings moisture absorption rate increases 1 time, moisture absorption rate of silicone and ParyleneC Coating is almost unchanged, ParyleneC moisture absorption rate is much less than other types of coatings. The effects of humidity test for the appearance of the coating and moisture absorption become more pronounced. In addition, the sort has a change in the rate of moisture absorption of the coating, acrylic coating of moisture absorption rate is greater than the silicone; salt fog test are quite the opposite by fungus test. This phenomenon indicates that acrylic coating has more effect on temperature, but organic silicon coating are sensitive to temperature and humidity changes than acrylic coating.

3.1.4 Data Analysis

(1) There is a good anti-mildew property for the four kinds of printed circuit board of coatings according to the results of fungus test.
(2) The moisture absorption of the four kinds of printed circuit board coating has been improved after the humidity. The main reason is as follows: as the temperature changes, condensation phenomenon will appear. It will prompt the coating to absorb moisture. What is more, the change of temperature can affect performance of the coating and lead to aging phenomenon. It can also promote moisture absorption of the coating.
(3) There is a little degeneracy phenomenon on the coating after salt fog test and humidity test by the appearance. But the insulation resistance is greater than 9,900 MΩ. It does not have a large impact on insulation performance.
(4) According to the analysis above, there is no obvious influence on the insulation resistance of the coating after corrosion test. Therefore, the second salt fog test and humidity test are put on. The test condition is the same with the first test in order to strengthen the impact on the coating by corrosion test.

3.2 The Second Test Results

Fungus test is got rid in the second test. The main reason is that fungus test has a very small influence on the coating added fungicide and the conditions are too mild, so it has little impact on appearance, moisture absorption and insulation resistance. Compared to the salt fog test and the humidity test, its influence can be neglected.

3.2.1 Salt Fog Test → Salt Fog Test

The second salt fog testis put on. The condition of the test is the same as the first salt fog test. The results are shown in Table 4:

From Table 4, we find all four of insulation resistances of coatings are more than 9,900 MΩ. It proved that after the second salt fog test, it still not to have a greater impact on the insulation resistance of the coating. For different types of coatings, moisture absorption rate from largest to smallest is: polyurethane coating > silicone coating > acrylic coating > parleneC coating. Compared to the result of the first salt fog test, moisture absorption rate of acrylic coating, polyurethane coating and silicone coating increased 100 %, but paryleneC coating moisture absorption rate was almost unchanged. The three kinds of coating could improve the moisture absorption, it shows that after the first test, moisture is not sufficient, moisture absorption rate of paryleneC coating remained almost unchanged, and the moisture absorption rate is kept to a minimum.

After the second salt fog test, we observed coatings by AFM. Surface morphology of the coating is shown in Figs. 2, 3, 4 and 5.

3.2.2 Humidity Test → Humidity Test

The second humidity is put on. The conditions of the test are the same with the first humidity test. The results are shown in Table 5.

From Table 5, we find that after the second humidity test, insulation resistance value of the coating is less than the minimum range (10 MΩ) in the edge of the solder paste and through-hole on printed circuit board. Obviously this part of the

Table 4 The results of salt fog test → salt fog test

Coating types	Moisture absorption rate (%)	Insulation resistance (MΩ)
Acrylic coating	1.1826	>9,900
Polyurethane coating	1.3769	>9,900
Silicone coating	1.0138	>9,900
ParyleneC coating	0.1211	>9,900

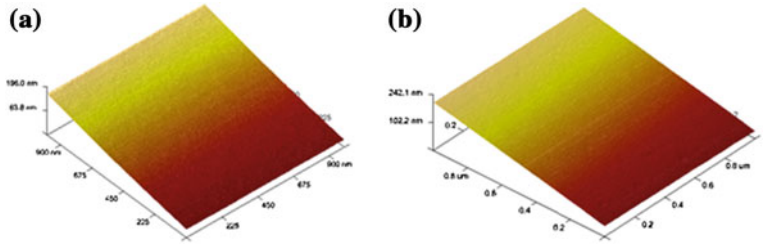


Fig. 2 Acrylic surface morphology by AFM. **a** The surface of the substrate, and **b** represents a solder paste components

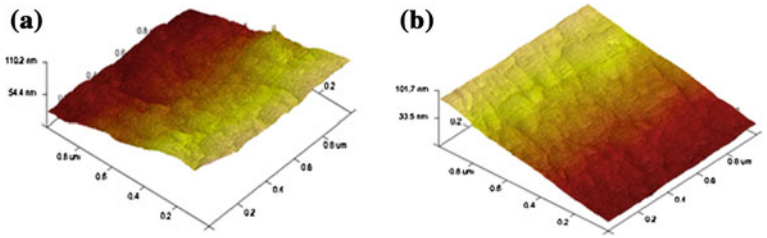


Fig. 3 Silicone surface morphology by AFM. **a** The surface of the substrate, and **b** represents a solder paste components

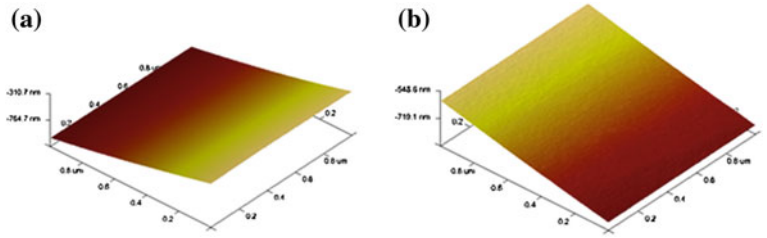


Fig. 4 Polyurethane surface morphology by AFM. **a** The surface of the substrate, and **b** represents a solder paste components

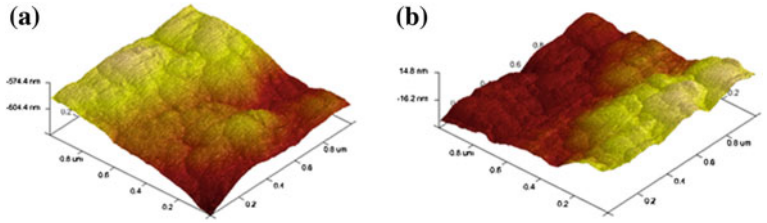


Fig. 5 ParyleneC surface morphology by AFM. **a** The surface of the substrate, and **b** represents a solder paste components

Table 5 The results of humidity test → humidity test

Coating types	Moisture absorption rate (%)	Insulation resistance (MΩ)
Acrylic coating	1.1005	a
Polyurethane coating	1.4187	>9,900
Silicone coating	0.97375	>9,900
ParyleneC coating	0.11755	>9,900

Note a insulation resistance value of the coating is less than the minimum range (10 MΩ) in the edge of the solder paste and through-hole on printed circuit board

coating is failure. Insulation resistance of other coatings is more than 9,900 MΩ, and insulation properties are good. Therefore speculated that acrylic coating are more sensitive to changes in temperature and humidity, Continuous data between 1 and 9,900 MΩ of insulation resistance is not detected. So the value of insulation resistance falls suddenly means the rapid failure of the coating. Moisture absorption rate compare to the first humidity increased about 10 %; the range of increases was smaller. It means that moisture absorption of coating in humidity test for the first time after almost reached saturation point. Four kinds of coating, moisture absorption rate of the sort generally is: Polyurethane coated > acrylic coated > silicone coating > ParleneC coating.

After two rounds of humidity test, we observed coatings using AFM. Surface morphology of the resulting photos part as shown in Figs. 6, 7, 8, 9.

3.2.3 Data Analysis

Compared the AFM photos after two rounds of humidity test with those after salt fog → salt fog test,we found that the morphology on the surface of the coating

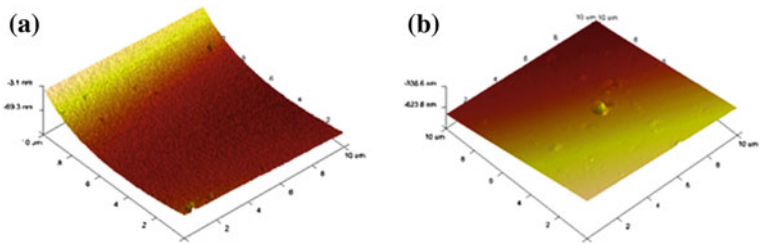


Fig. 6 Acrylic surface morphology of AFM. **a** The surface of the substrate, and **b** represents a solder paste components

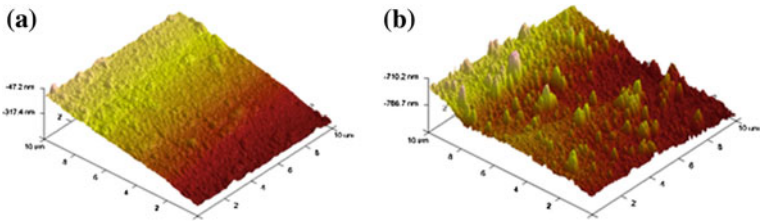


Fig. 7 Silicone surface morphology of AFM. **a** The surface of the substrate, and **b** representsa solder paste components

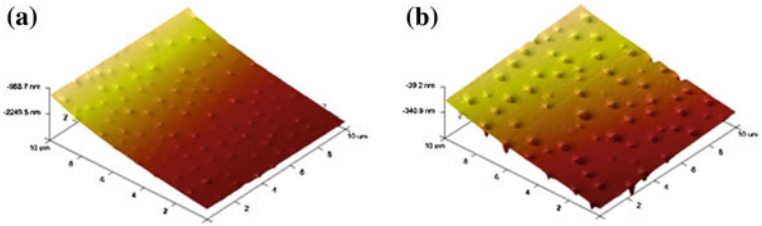


Fig. 8 Polyurethane surface morphology of AFM. **a** The surface of the substrate, and **b** represents a solder paste components

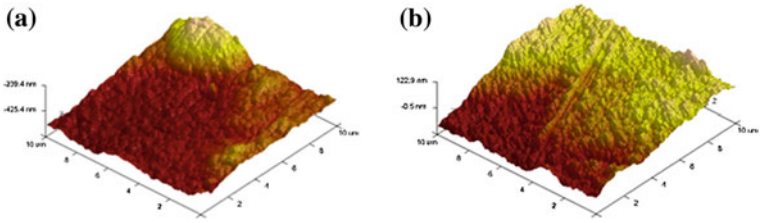


Fig. 9 ParyleneC surface morphology of AFM. **a** The surface of the substrate, and **b** represents a solder paste components

changes larger and the coating bump is more serious after humidity test → humidity test. Therefore, we speculated that changes in temperature and humidity have a more significant effect on organic coatings.

4 Conclusions

(1) After fungus test, humidity test, salt fog test on the four coatings, we found that humidity test play the most important role on the rate of moisture absorption, then salt fog test, the last fungus test during the corrosion tests.

And in terms of moisture absorption rate, humidity is equal to two round of salt fog test.

(2) Addition to moisture absorption rate of the ParyleneC printed circuit board coating is relatively small, saturated moisture absorption rate of the remaining three are over 1 %.

(3) According to moisture absorption rate, appearance of coating after corrosion test and insulation resistance, these three properties of coatings from superior to inferior sorts are: ParleneC coating > Silicone coated > Polyurethane coating > Acrylic coating. The main reasons are as follows: Acrylic coating after a humidity or salt fog test, acrylic coating appears to change color, powder and so on. After humidity test → humidity test, it appears insulation failure, and insulation resistance reflects the barrier properties of the coating on the media. So the performance is the worst. Polyurethane coating in the first round after the humidity test and salt fog test, came up the discoloration, and its moisture absorption rate was always the biggest in several coatings, planted which will actually cause a larger risk, but the insulation failure does not occur throughout the process, which means the protective properties of coatings have not lost, So corrosion performance is slightly better than the acrylic coating. ParleneC coating moisture absorption rate is minimum, moisture proof is the best and the insulation performance is also good, and it did not appear clearly chalking, color changing throughout the test procedure. Therefore, its performance is superior to the other three kinds. Silicone coating's overall performance is second only to the ParleneC coating, but superior to polyurethane and acrylic coatings.

(4) From the results of the experiment we can also conclude that the edge of the solder paste and through-hole on the coating are weaknesses of the acrylic coatings. Continuous data between 1 and 9,900 MΩ of insulation resistance is not detected. So the value of insulation resistance falls suddenly means the rapid failure of the coating. If we can discover the dump earlier, then its failure life can be predicted earlier, which is a direction of our research on this project in the future.

(5) This paper provides many detection methods about the coating on Printed circuit board. The weakness such as the edge of the solder paste and through-hole of the printed circuit board are confirmed.

References

1. Funke W (1997) Problems and progress in organic coatings science and technology. Prog Org Coat 31(1):5–9
2. Merlatti C, Perrin FX, Aragon E, Margaillan A (2008) Natural and artificial weathering characteristics of stabilized acrylic–urethane paints. Polym Degrad Stab 93(5):896–903
3. Zhang YY, Deng HX, Shi HJ, Yu HC, Zhong B (2012) Failure characteristics and life prediction for thermally cycled thermal barrier coatings. Surf Coat Technol 206 (11):2977–2985

4. Bacos MP, Thomas M, Raviart JL, Morel A, Mercier S, Josso P (2011) Influence of an oxidation protective coating upon hot corrosion and mechanical behaviour of Ti–48Al–2Cr–2Nb alloy. Intermetallics 19(8):1120–1129
5. Qian M, Mcintosh Soutar A, Tan XH, Zeng XT, Wijesinghe SL (2009) Two-part epoxy-siloxane hybrid corrosion protection coatings for carbon steel. Thin Solid Films 517 (17):5237–5242
6. Bierwagen GP (1996) Reflections on corrosion control by organic coatings. Prog Org Coat 28 (1):43–48
7. GJB150.9 Environmental test methods for military equipments-Damp heat test. Commission on Science, Technology, and Industry for National Defense, China
8. GJB150.10 Environmental test methods for military equipments-Fungus test. Commission on Science, Technology, and Industry for National Defense, China
9. GJB150.11 Environmental test methods for military equipments-Salt fog test. Commission on Science, Technology, and Industry for National Defense, China
10. QJ519A-1999 Test methods for printed circuit board. Aerospace Science and Technology Corporation, China
11. GB/T 1766-2008 Paints and varnishes-Rating schemes of degradation of coats. Standardization Administration of the People's Republic of China

Improvements in Computed Order Tracking for Rotating Machinery Fault Diagnosis

K.S. Wang, D.S. Luo, W. Guo and P.S. Heyns

Abstract This paper introduces two improved procedures in computed order tracking analysis which enhances the ability of the method for rotating machinery fault diagnosis. Firstly, the method for the improvement of resolution of the resultant order spectrum is introduced. Secondly, an alternative method to computed order tracking analysis which avoids time to angle domain transformation for measured data is discussed. The two improvements enhance abilities to computed order tracking analysis in terms of better order spectrum resolution and alternative implementation procedure to obtain order spectrum of the signals. The proposed improvements to order tracking analysis are useful for rotating machine condition monitoring and are expected for further applications in real practices.

1 Introduction

Computed order tracking (COT) is, nowadays, one of the widely used and researched order tracking techniques. It excludes speed variation effects in measured data by re-sampling time domain signals to angle domain so that the subsequent Fourier analysis may present clear spectrum components [1]. Researchers have extensively studied the method in its theory and applications [1–3]. However, due to assumptions and constraints in the procedures of COT, several unfavourable restrictions exist in some implementations. Such restrictions include for instance, the constant angular acceleration assumption which artificially assumes a linear rotational speed in a revolution to re-sample time signals to the angle domain; the use of polynomial

K.S. Wang (✉) · D.S. Luo · W. Guo
School of Mechanical Electronic and Industrial Engineering, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China
e-mail: keshengwang@uestc.edu.cn

P.S. Heyns
Department of Mechanical and Aeronautical Engineering, Dynamic Systems Group, University of Pretoria, Pretoria 0002, South Africa

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_116

interpolation to find re-sampled data amplitudes which violate the sinusoidal cyclic nature of rotating machine vibrations; fewer revolutions of the measured data that limits the order resolution in the final order spectrum, etc. [1]. As a result these procedures influence the accuracy of the resultant order spectrum and compromise the diagnostic ability of COT. However for the sake of simplicity of the method, these disadvantages are accepted and are to a large extent unavoidable. Simple methods to avoid and improve the consequences of the above limitations of COT are beneficial to the diagnostic ability of the method.

In this paper, firstly a zero padding method which was originally proposed to the DFT for a better approximation of Fourier transform is briefly introduced. However, here the zero padding method is discussed in terms of angle domain data instead of original time domain data. Advantages of the zero padding method are considered and further developed into a non-zero padding method which is specially for rotating machine vibration data. The introduction of non-zero padding method in COT enhances the ability of COT in machine fault diagnostics. Secondly, an approach which uses measured signal and rotational speed based complex exponential functions to calculate modified Fourier coefficients to obtain order spectra, is proposed. It could be an alternative way to COT method with absence of data re-sampling procedures and avoids most of the limitations and constraints from traditional COT method. In the following, the first section briefly explains the basic rationale of those methods and then the following section demonstrates the abilities of the proposed methods through some synthetic signals.

2 Basic Rationale

2.1 Zero Padding and Non-zero Padding Method

Basically, zero padding is to add a string of zeros to a signal so that the frequency step by DFT can be decreased therefore a better frequency resolution spectrum can be obtained [4]. "The zero padded data may be thought of as an infinite signal multiplied by a finite length rectangular window in time domain and this process is equivalent to introduce a convolution of the signal with a sinc function" (where sinc = sin(x)/x) in the frequency domain. The main purpose of the technique is to allow a better approximation to the Fourier transform of a signal to be calculated using DFT [4]. For order tracking analysis, the zero padding method may be introduced into the analysis of angle domain signals which is transformed from measured time domain data in terms of rotational speed, e.g. COT. Naturally, the introduction of zero padding may decrease the order step of resultant COT spectrum and therefore, improves the quality of resultant order spectrum. The details of theoretic background are discussed by author's previous report, readers may refer to [5].

However, it should be realized that although the order step has been decreased by zero padding, the signal that is being analyzed remains the same. In other words, a rectangular window applied on the zero padded signals yield the same result of

the measured signals and for this reason, executing a DFT in the order domain does not really represent a true improvement in order resolution but rather a better approximation of the order spectrum through the zero padding method. This however leads to the further thought that if more data are being analyzed and it is equivalent to a wider rectangular window applied in angle domain and therefore results in a narrower sinc function in order domain, an improvement of order resolution may result. Thus, a simple assumption can be made as the measured rotating machine vibration data repeats itself for several other revolutions. It is reasonable to make this assumption that the rotating machine vibrations are cyclicly repeating every revolution. If such an assumption is made, the angle domain data can be non-zero padded. In short, for COT, angle domain signal could be modified in terms of zero or non-zero padding methods. Order resolutions in the order spectrum could be improved. A comparison of traditional computed order tracking (COT), zero and non-zero padding COTs is made in Table 1 [5].

2.2 *Order Tracking Analysis from Rotational Speed Based on Fourier Coefficient Calculations*

COT has been widely used over decades, however researchers have also pointed out disadvantages of the method, e.g. that it involves several artificial assumptions and data manipulations which inevitably introduce errors and compromise diagnostic ability of the method [1, 5]. Unfortunately, nowadays there is almost no equivalent method to COT for order tracking analysis and the analyst usually has to live with the limitations of the method. In this paper, a rotational speed based Fourier coefficient calculations is proposed through which an alternative way to obtain order spectrum in terms of Fourier calculations. And by doing so, data re-sampling procedure is avoided and therefore most of the limitations and constrains from traditional COT method are eliminated. In the following, the basic rationale of method will be discussed.

It is well known that simple sinusoids are commonly considered as the building blocks of most complicated signals. Generally speaking, the methods to determine those building blocks are usually referred to as Fourier analysis. Considering rotating machine vibration signals, the typical characteristics are its rotational speed frequencies. And in most cases, the rotational speed related vibrations comprise most of the energy in the measured data. Therefore, it is reasonable to rebuild signals by special building blocks, i.e. rotational speed vibrations and their harmonics. By doing so, the rotating machine vibrations, in fact, are interpreted in terms of rotational speed and a perspective in terms of order(s) or rotating speed is provided. Notice that, this process does not involve procedures to transform signals from the time to the angle domain which is inevitable in traditional COT. Therefore, it offers a possible solution to order tracking analysis without introducing artificial errors. To briefly introduce its theoretical basis, the logic of above method could be depicted as in Table 2.

Table 1 Comparison on COT, zero padding COT and non-zero padding COT

	COT	Zero padding COT	Non-zero padding COT
Data Analyzed	x_n	x_n	x_n
	$n = -N/2 + 1 \rightarrow N/2 + 1$	$n = -\alpha N/2 + 1 \rightarrow \alpha N/2 + 1$ *if* $n \notin -N/2 + 1/2 + 1,\ x_n = 0$	$m = -\alpha N/2 + 1 \rightarrow \alpha N/2 + 1$ $x_m \Leftrightarrow \underbrace{x_1, \ldots x_n, x_1, \ldots x_n, \ldots \ldots, x_n}_{\alpha N}$
Order step	o	o/α	o/α
Equivalent rectangular window length	N	N	αN

where, N—the number of samples of original data
αN—corresponds to α times original measured revolutions for the angle domain data

Table 2 Rationale of the proposed method

$x(t)$ decomposition → in terms of traditional Fourier Series representation	$x(t) = \sum_{k=0}^{+\infty} a_k e^{2\pi jkt}$ →	Frequency spectrum a_k at complex exponential with fundamental frequency k
⇓		
Modify complex exponential by rotational speed frequency $f(t)$		
⇓		
$x(t)$ decomposition → in terms of the proposed rotational speed based representation	$x(t) = \sum_{m=0}^{+\infty} a_m e^{2\pi jmf(t)}$ →	Frequency spectrum a_m at complex exponential of rotational speed frequency, $f(t)$

where $x(t)$—a general form of any rotating machine vibrations
a_k—Fourier coefficient
a_m—Modified Fourier coefficient
$f(t)$—Rotational speed frequency

According to Table 2, for each k, the stationary complex exponential, $e^{2\pi jkt}$, is then modified to be possible non-stationary in terms of rotational speed, $e^{2\pi jmf(t)}$. Further, based upon the theory of Fourier analysis, the term a_m could be calculated by a dot product between $x(t)$ and $e^{2\pi jmf(t)}$, i.e. $a_m = \int_0^1 e^{-2\pi jmf(t)} x(t) dt$. Once a_m is determined, the order spectrum results. This process, in fact, is to represent the signal $x(t)$ through combination of possible non-stationary complex exponentials in terms of rotational speed $f(t)$. By doing so, the modified Fourier coefficient a_m or order spectrum of $x(t)$ is brought out.

3 Demonstration Through Synthetic Signal

3.1 Demonstrations for Zero and Non-zero Padding Methods

A 10 s analytical signal consisting of a sum of three orders, that is 3rd, 3.2 th and 5 th, is considered for demonstration of non-zero padding method. A varying rotational speed is simulated through which the rotor complete two revolutions within 10 s. The details of the simulation signals are listed in Table 3. Readers may also refer to Ref. [5].

Traditional computed order tracking (COT) is applied to the simulated signal y, the time waveform is firstly transformed into evenly distributed angle domain waveforms. Based upon this angle domain signal, further Fourier analysis which follows the traditional COT, zero padded COT and non-zero padded COT are performed and order spectral are depicted as in Fig. 1.

Table 3 Signal simulation

Initial conditions	Sampling rate	Angles that rotor turns through (θ)	Angular acceleration (α)	Angular speed (ω)
$t_0 = 0; \omega_0 = 0$ $\alpha_0 = (8\pi)/100$	Time domain: 1,024 points/ Second. Angle domain: 500 points/ revolution	$\theta = \omega_0 t + \frac{1}{2}\alpha t^2$	$\alpha = (8\pi)/100$ α is determined by substituting $\theta = 2 \cdot 2\pi t = 10s$ into $\theta = \omega_0 t + \frac{1}{2}\alpha t^2$	$\omega = \alpha t$
Simulated analytical signals	$y = \sin(3\omega t) + \sin(3.2\omega t) + \sin(5\omega t)$			

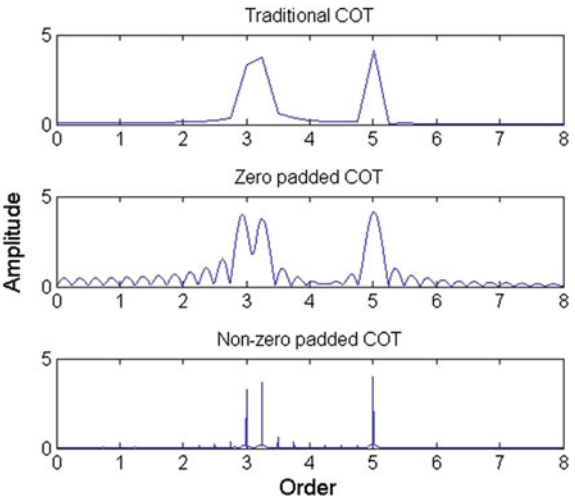


Fig. 1 COT, zero and non-zero padded COT

As can be expected, the order step of traditional COT is

$$\Delta o = \frac{1}{R} = \frac{1}{2} = 0.5\,(order).$$

where R is revolutions of the measured signal. This is, of course, is not good enough for the close orders, 3rd and 3.2th order. It can be seen in the top figure of Fig. 1 that 3rd and 3.2 th orders are mixed together within one thick order peak. The COT cannot distinguish them. Then the zero padding method is applied to the re-sampled angle domain signals and the new order step becomes

$$\Delta o = \frac{1}{R} = \frac{1}{40} = 0.025\,(order).$$

The order spectrum is shown in the middle figure of Fig. 1. It is good to see that two close round order peaks emerge after the zero padding. However it should also be noticed that quite a few side lobes of order peaks are obvious which may possibly lead the confusions to other order peaks or sidebands. This effect is actually because the equivalent length of the rectangular window does not increase with the increase of data length. And therefore the corresponding sinc function in order domain cannot be narrowed which produces several side lobes in the order spectrum. Further, the non-zero padding method is applied to the re-sampled angle domain signals to overcome this drawback, as is shown in the bottom figure of Fig. 1. Clearly, three distinct order peaks are rendered, there are no more side lobes and narrow order peaks are obtained. This is the result of the increase of revolutions which decrease the order step and the increase the width of rectangular window which narrowed the corresponding sinc function peak in order domain.

In short, zero and non-zero padded COT overcomes the limitations of traditional COT in order spectrum resolution and non-zero padded COT further overcomes the shortcoming of extra side lobs for the zero padded COT in order spectrum.

3.2 Demonstrations for Order Tracking Analysis from Rotational Speed Based Fourier Coefficient Calculations

In order to validate the proposed method, vibration signal from a single-degree-of-freedom rotor simulation model is used for demonstration. The lateral response of a symmetric rotor is modelled as two uncoupled single degree of freedom systems. It is assumed that a rotor of mass m is mounted on bearings of total stiffness k and damping coefficient c. The rotor rotates at an increasing rotational speed to

Table 4 Signal Simulation

Parameter	Value
Stiffness k	500,000 N/m
Rotor mass m	20 kg
Eccentricity r_u	0.1 m
Unbalance mass m_u	0.05 kg
Damping coefficient c	100 Ns/m
Initial time t_0	0 s
Final time t_f	5 s
Number of time steps	4,096
Number of revolutions	100
Angular speed ω_1	$\omega = 15.0796\ t^2$ rad/s
External excitation	$F = A\sin(5\omega t) + 0.8A\sin(5.1\omega t)$, $A = m_u\omega^2 r_u$

complete 100 revolutions within 5 s. For details of the model, readers may refer to Wang and Heyns [6]. However, different external excitation models in this simulation study are adopted to explicitly demonstrate the ability of the proposed method. The characteristics of the model are shown in Table 4, considering the response in the vertical direction.

From Table 4, one can envision some distinct features of the simulated signals. Firstly, mass m and stiffness k are two important parameters that determine the resonance frequency of the system which is proportional to the square root of k/m. Secondly, the vibration amplitude and frequency of simulated signals are not constant but varying with rotational speed for the external excitation is being set. Thirdly, two close orders, i.e. 5 th and 5.1 th are rendered in the system responses due to the two closely spaced external excitations, defined by $F = A_m \sin(5\omega t) + 0.8A_m \sin(5.1\omega t)$. The nature of simulated signals is discussed and the proposed method demonstrated in the following.

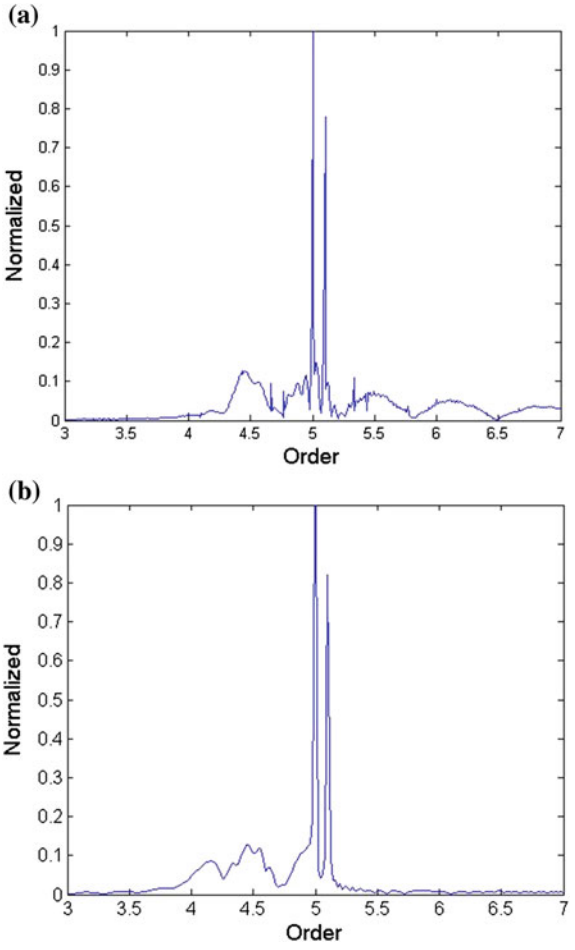


Fig. 2 COT and the proposed method. **a** Computed order tracking (COT) number of sampling intervals = 100, **b** Modified Fourier coefficient based order tracking

Since the proposed method yields order spectra which are similar to those obtained from computed order tracking (COT), is first applied and shown in Fig. 2a for comparison. In order to make the order spectrum maps more comparable, the order spectra are all normalized in terms of the maximum value of each spectrum map.

Figure 2a shows that the 5 th and 5.1 th spectra are explicitly differentiated by COT in terms of the order spectrum. Then, the proposed method is applied to the simulated signals and order spectrum computed through the proposed method is presented in Fig. 2b where orders from 3 to 7 with a step size of 0.05 orders are chosen. That is $m = 3, 3.05, 3.1, 3.15, \ldots, 7$. It is encouraging to see that the 5 th and 5.1 th orders are again explicitly differentiated and two figures in Fig. 2 are nearly identical. This comparison demonstrates that the proposed method yields similar results to the conventional COT method and suggests that the method could be a substitution to COT for order spectrum analysis. However, it should be borne in mind that, the proposed method, is a totally different method from COT with the absence of signal transformation from time to angle domain and most of limitations are naturally avoided.

4 Conclusions

In this paper, the two improvements on computed order tracking (COT) are discussed. Firstly, zero and non-zero padding methods are introduced to overcome the limitations of order spectrum resolution. The results show that the zero and non-zero padded COT can both effectively decrease the order step of order spectrum which is useful for distinguish close orders. Further, a modified Fourier coefficient based order tracking is proposed. It can be a substitution to COT analysis. Compared to COT, the proposed method yields similar order spectrum and with absence of data re-sampling process, to a large extent, minimizes artificial errors and enhances diagnostic ability to order tracking analysis. Two improvements of order tracking analysis are both useful and effective in the augment of COT in vibration monitoring. Further research on the theory as well as experimental demonstrations are currently underway.

Acknowledgment The work that is described in this paper is fully supported by the Fundamental Research Funds for the Central Universities (Project No. ZYGX2012J105). We appreciate three anonymous reviewers for their valuable comments and suggestions for improving this paper.

References

1. Blough JR (2003) A survey of DSP methods for rotating machinery analysis, what is needed, what is available. J Sound Vib 262(3):707–720
2. Fyfe KR, Munck EDS (1997) Analysis of computed order tracking. Mech Syst Signal Process 11(2):187–2059

3. Eggers BL, Heyns PS, Stander CJ (2007) Using computed order tracking to detect gear condition aboard a dragline. J South Afr Inst Min Metall 107(2):115–122
4. Neild SA, McFadden PD, Williams MS (2003) A review of time-frequency methods for structural vibration analysis. Eng Struct 25(6):713–728
5. Wang KS, Heyns PS (2012) A non-zero padding method in the angle domain to improve the order spectrum resolution of computed order tracking. In: Proceedings of condition monitoring of machinery in non-stationary operations, part 3, pp 375–383
6. Wang KS, Heyns PS (2009) Vold-Kalman filter order tracking in vibration monitoring of electrical machines. J Vib Control 15(9):1325–1347

Application of Reliability Growth Model in Step-Down Stress Accelerated Storage Test

YaHui Wang, XiaoGang Li and TaiChun Qin

Abstract Step-down accelerated storage test is equalled to reliability growth, a method of step-down stress accelerated storage test based Army Materiel System Analysis Activity (AMSAA) model is proposed, according to the step-down stress, the mean time between failures (MTBF) under normal stress can be obtained, by using the AMSAA model. First, through the product's cumulative failure time, cumulative failure numbers and based on AMSAA model, the product's instantaneous MTBF is shown. Then the step-down stress is divided into n ladders, and there is only one fault under each stress, the joint probability density function is shown by the cumulative failure time of each step-down stress, and the parameters of the AMSAA model are estimated by the maximum likelihood estimation, then the point estimation of the product's instantaneous MTBF is got. By choosing instantaneous MTBF of certain accelerated stresses and combined respective physical accelerated model, the product's storage lifetime is estimated. Finally, a case study is performed using this method. The effectiveness of this method is shown, for the point estimation of each parameter is little different. Thus it provides a new evaluation method for step-down accelerated storage test.

Keywords Step-down stress · Reliability growth · AMSAA model · MTBF · Physical accelerated model

1 Introduction

As for missile products, it also has an important characteristic besides use function, that is, it needs the long-term storage before using, therefore carrying out storage test of missiles has a strategic significance. The assessment of storage reliability is very difficult under normal stress level; however, the problem can be solved well by

Y. Wang · X. Li (✉) · T. Qin
School of Reliability and System Engineering, Beihang University, Beijing, China
e-mail: lxg@buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_117

the accelerated life testing (ALT). Based on the differences of the stress loaded, ALT can be divided into as: constant stress accelerated life testing, step-stress accelerated life testing, and progressive stress accelerated life testing. At present, among the accelerated life testing, the application of constant stress accelerated life testing is comparatively mature. But tests at constant,the required samples are comparatively large and the testing time is also relatively long [1], on the contrary, step-stress accelerated life testing conquers these problems, and gets more preference in the engineering. During step-stress accelerated life testing, there is little fault under initial low stress, thus the initial stress and test time are not easily controlled.

Considered above problems, Zhang and Chen [2, 3] proposed a method of step-down accelerated life testing, and studied the effectiveness by theoretical model, Monte Carlo simulations, and comparative experiments. Wang and Zhang [4] proposed a new optimizing design method of step-down accelerated life testing based on Monte Carlo simulations. Tan [5] built the data conversion formula of step-down accelerated life testing by accelerated model, and then made improvement for the three step method of step-down test. Xu [6] compared the effectiveness between step-down test and step-stress test by Monte Carlo simulations, and obtained the condition when the efficiency of the step-down test was superior to that of the step-stress test.

According to the thought of reliability growth, this paper puts forward a method of step-down accelerated storage testing based on Army Materiel System Analysis Activity (AMSAA) model, which divides step-down stress into n ladders; sets there is only one failure under each stress; describes the joint probability density function by the cumulative failure time of each step-down stress; and estimates the parameters of the AMSAA model by the maximum likelihood estimation to obtain the point estimation of the product's instantaneous MTBF. We combine instantaneous MTBF of certain accelerated stresses with relative physical model to estimate the storage lifetime of products. This paper selects temperature as accelerated storage stress and Arrhenius model as accelerated storage model.

2 AMSAA Model in the Step-Down Accelerated Storage Testing

2.1 Calculate the Instantaneous MTBF of Products by AMSAA Model

When we exert step-down stress on products, at the beginning, there will be a lot of failures, with the reduction of stress level, the fault will gradually decrease. Obviously, more fault information can be obtained by the step-down accelerated stress testing. We can use Reliability Enhancement Test (RET) to ascertain the highest stress level which can load on products [7], as long as the initial stress is

proper, the failure mechanism will not be changed. In this experiment, the cumulative number of failures is a stochastic process, which exactly corresponds with the theory of AMSAA model; According to AMSAA model, the cumulative number of failures is an inhomogeneous Poisson distribution.

When we performance the step-down accelerated storage testing for repairable electronic products, according to the relation between the cumulative number of failures and the accumulated accelerated storage time, the instantaneous failure rate of products is given by [8–11]:

$$\lambda(t) = abt^{b-1} \tag{1}$$

where a > 0 is a scale parameter, b > 0 is a shape parameter. The instantaneous MTBF of products is the inverse of *instantaneous* failure rate, which is shown as follows:

$$M(t) = \frac{1}{\lambda(t)} = \frac{1}{abt^{b-1}} = \left[abt^{b-1}\right]^{-1} \tag{2}$$

2.2 Statistical Analysis of AMSAA Model

Due to AMSAA model is established based on rigid stochastic process theory, in the step-down accelerated storage lifetime testing, when we use AMSAA model to estimate the storage lifetime of products, there exists a series of statistical analysis methods, that is, the censored time and the censored data, and this paper chooses the censored data.

Suppose that the number of failures at each stress condition is $N_{S_1}, N_{S_2}, \ldots, N_{S_n}$, and its corresponding failure time is $t_{S_1}, t_{S_2}, \ldots, t_{S_n}$. Then we can get the accumulated failure is $t_j = \sum_{i=1}^{j} t_{S_i}$, and the accumulated number of failures is $N(t_j) = \sum_{i=1}^{j} N_{S_i}$.

Hence, the likelihood function of the data of the censored data can be obtained by:

$$\mathrm{f}(\mathrm{t_{S_1}}, \mathrm{t_{S_2}}, \ldots, \mathrm{t_{S_n}}) = (\mathrm{ab})^{\mathrm{N(t_n)}} \mathrm{e}^{-\mathrm{at_n^b}} \times \prod_{\mathrm{j=1}}^{\mathrm{n}} \mathrm{t_j^{b-1}} \tag{3}$$

Then in the AMSAA the maximum likelihood estimation of (a, b) is given by:

$$\widehat{\mathrm{b}} = \frac{\mathrm{N(t_n)}}{\sum_{\mathrm{j=1}}^{\mathrm{n}} \ln \frac{\mathrm{t_n}}{\mathrm{t_j}}} \tag{4}$$

$$\hat{a} = \frac{\mathrm{N(t_n)}}{\mathrm{t_n^{\hat{b}}}} \tag{5}$$

Thus when the step-down stress is down to relative accelerated stress, the instantaneous MTBFi can be obtained, and the maximum likelihood estimation of MTBF is given by:

$$\hat{\theta}(t_j) = \left[\hat{a}\hat{b}t_j^{\hat{b}-1}\right]^{-1} \tag{6}$$

3 The Computation of Storage Lifetime by Accelerated Model

This paper selects temperature as accelerated storage stress, thus the accelerated model is Arrhenius model, which can be written as:

$$\xi = Ae^{E_a/kT} \tag{7}$$

where ξ denotes some life characteristic, A is a constant, k denotes Boltzmann's constant, as $8.617 \times 10^{-5} eV/K, E_a$ denotes activation energy. From Eq. (7), with the increment of temperature, the life characteristic decreases exponentially. By taking the natural logarithm of both sides of Eq. (7), we can get:

$$\ln \xi = d + e/T \tag{8}$$

Based on Eq. (6), the instantaneous MTBF can be obtained at each temperature stress, in the step-down accelerated storage lifetime, we can choose some step-down accelerated stress, then combine the step-down accelerated stress with corresponding instantaneous $MTBF_i$, and get several groups of value of temperature and lifetime. After that, the least-squares estimator (LSE) $\hat{d}, \hat{e}$ of d, e can be obtained based on Eq. (8), and the storage lifetime under normal stress can be obtained.

4 An Illustrative Example

Assuming that assembly products' lifetime is an exponential distribution, $\theta = 2,000$h, temperature as the accelerated stress, the step-down stress accelerated test can be considered as reliability growth test, which the first step data T_1, that is the highest stress which is called the initial value of growth test, can be ascertained based on RET. This paper assumes that $T_1 = 85$ °C, when there exists only one failure, the test will be turned to the next temperature ladder, we take 2.5 °C as a ladder. Based on the above conditions, the simulation failure data is listed in Table 1.

Based on the Eqs. (4) and (5), the MLE $\hat{a}$ and $\hat{b}$ can be computed, then we can count the instantaneous MTBFi respectively from the highest stress condition 85 °C to accelerated stress condition 35, 40, 45, 50 °C, which are listed in Table 2.

Table 1 The cumulative failure data of assemblies

N	N (t_j)	T_i (k)	T_j (h)
1	1	358	4.4
2	2	355.5	10.7
3	3	353	16.4
4	4	350.5	22.5
5	5	348	33.9
6	6	345.5	46.4
7	7	343	59.6
8	8	340.5	79
9	9	338	94
10	10	335.5	132.9
11	11	333	165.8
12	12	330.5	211.2
13	13	328	297.2
14	14	325.5	369.1
15	15	323	473.8
16	16	320.5	607.7
17	17	318	866.2
18	18	315.5	1127.3
N	**N (t_j)**	**T_i (k)**	**T_j (h)**
19	19	313	1393.6
20	20	310.5	1743.2
21	21	308	2553.8

Table 2 The MTBF of the different temperatures

Stress (T_i)	35 °C	40 °C	45 °C	50 °C
a_i	1.58	1.183	1.51	0.9032
b_i	0.3233	0.3835	0.3891	0.4447
t_j	2553.8	1393.0	866.2	473.8
$MTBF_i$	566 h	192 h	106.4 h	76.9 h

From the Table 2 we can get, when the highest stress growths to the different stress, the difference of each parameter is small. Thus it illustrates that the growth rate is considered as constant in the step-down stress reliability growth test, that is, we use the thought of reliability growth to carry out the step-down stress accelerated storage test is feasible.

Based on the Table 2, the relation between the instantaneous $MTBF_i$ under accelerated stress and corresponding accelerated stress is shown in the Fig. 1.

Based on the Eq. (8), the LSE of *d, e* can be obtained, that is, $\hat{d} = -36.85$, $\hat{e} = 13250$. Then the storage lifetime under normal stress can be extrapolated.

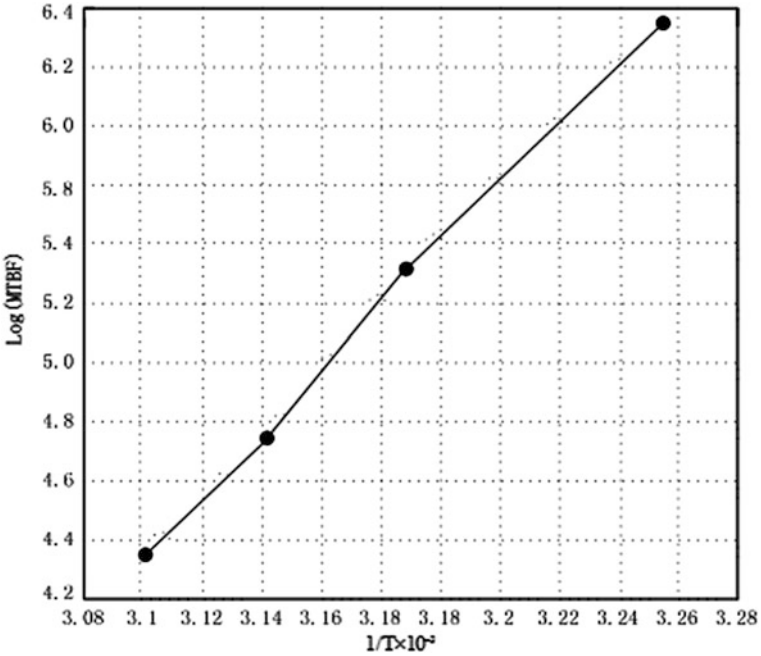


Fig. 1 The fitting chart between life and stress

$$\theta_0 = \exp(-36.85 + 13250/(273 + 25)) = 1998.2\text{h} \tag{9}$$

The result is nearly the same with real MTBF 2000 h of the assembly products.

5 Conclusion

This paper presents a method which uses the reliability growth model in the step-down stress accelerated storage test, which avoids the trouble of data conversion. Then we compute an example, the result of parameter estimation illustrates that, when the highest stress growths to different stress, the growth rate is constant, which correspond with the thought of reliability growth. Finally, the storage lifetime of the assembly products can be obtained by accelerated model, which is nearly the same with real MTBF. Thus it provides a new evaluation method for step-down accelerated storage test.

References

1. Nelson W (1980) Member ASQC. Accelerated life testing-step-stress models and data analysis. IEEE Trans Reliab 63(2):103–108
2. Zhang CH, Chen X, Wen XS (2005) Step-down-stress Accelerated Life Testing-Methodology. Acta Armamentaril 26(5):661–665
3. Zhang CH, Chen X, Wen XS (2005) Step-down-stress Accelerated Life Testing-Statistical analysis. Acta Armamentaril 26(5):666–669
4. Wang YS, Zhang CH, Chen X (2007) Step-down-stress accelerated life testing-optimal design. Acta Armamentaril 28(6):686–690
5. Tan W, Shun YM, Sun YD (2011) Reliability simulation for step-down-stress accelerated life testing. Computer Simulation 28(12):80–83
6. Xu G, Wang RH (2008) Efficiency analysis of the step-down-stress accelerated life testing. J Shanghai Norm Univ 37(5):468–475
7. Ye DZ (2000) Reliability enhancement test. In: Tenth annual conference of the Chinese institute of electronics reliability branch, pp 40–48
8. IEC/TC56 (1989) Draft, reliability growth models and estimation method
9. Green JE (1973) The problems of reliability growth and demonstration with military electronics. Microelectr Reliab 12:513–520
10. GJB1407-92 (1992) Reliability growth test
11. Crow L H. AMSAA reliability growth symposium. ADA027053, 1974
12. Zhao Y, Yang J, Ma XB (2009) Data analysis of reliability. Beihang University Press, Beijing, pp 234–235
13. He GW, Dai CZ (2003) Reliability testing technique. National Defence Industry Press, Beijing, pp 159–163
14. Mao SS, Wang LL (1995) Accelerated life test. Science Press, Beijing, pp 18–19

Wireless Condition Monitoring Integrating Smart Computing and Optical Sensor Technologies

Christos Emmanouilidis and Christos Riziotis

Abstract Condition monitoring is increasingly benefitting from the application of emerging technologies, such as mobile computing and wireless sensors, including photonics sensors. The latter can be applicable to diverse application needs, due to their versatility, low costs, installation and operational flexibility, as well as unique safety and reliable operation characteristics in real industrial environments of excessive electromagnetic interference and noise. Coupling the monitoring flexibility offered by photonics technologies, with the data transmission flexibility of wireless networking provides opportunities to develop hybrid wireless sensor solutions, incorporating optical sensors into wireless condition monitoring architectures. This paper presents ongoing work within an integrated architecture for condition monitoring and maintenance management support, exploiting the added value of optical technology, inherently safe with respect to electromagnetic compatibility. The reported results are part of a collaborative project involving technology providers in wireless sensor networking, embedded systems and maintenance engineering, as well as research organizations active on photonics technologies and informatics for wireless and intelligence-enabled engineering asset management. The industrial test cases are from a lifts manufacturing industry, focusing on both production facilities assets, as well as on the end-product. The photonic platform of plastic optical fibers was selected due to its versatility and suitability for rapid customization and prototyping. The platform can serve diverse sensing and monitoring needs, ranging from physical parameters as strain and displacement in machinery parts, to chemical and biochemical monitoring of industrial-grade coolants' aging. Use of novel nanostructured optical materials together with laser-based micromachining techniques enabled the functional enhancement through rapid prototyping of optical fiber devices towards highly

C. Emmanouilidis (✉)
ATHENA Research & Innovation Centre, Athens, Greece
e-mail: chrisem@ceti.athena-innovation.gr; christosem@ieee.org

C. Riziotis
National Hellenic Research Foundation, Athens, Greece
e-mail: riziotis@eie.gr; riziotis@gmail.com; Christos.Riziotis@ieee.org

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_118

customizable sensors. The integration of the sensing elements within the wireless sensor network architecture offers substantial flexibility for industrial applications.

Keywords Photonics · Optical fibers · Sensors · Wireless sensors · E-maintenance

1 Introduction

The incorporation of emerging information and communications technologies (ICT) and photonics-based sensing in industrial and infrastructure monitoring holds considerable potential that is still not sufficiently well explored in Physical Asset Management [1]. Among the constituents of enabling ICT technologies are wireless, web-based, semantic and context-adaptive computing coupled with machine learning [2, 3], whereas wireless sensing [4] and photonic technologies [5, 6] offer increased flexibility and versatility on the sensing and signal transmission side. The integration of such enabling ICT and photonic technologies within an e-maintenance architecture seeks to take advantage of the aforementioned benefits while pushing towards greater data and services integration in physical asset management [7].

This chapter looks into the potential of photonics technologies for condition monitoring, with a particular focus on the integration of wireless sensing, mobile and adaptive computing and devices for implementing flexible smart monitoring solutions. The prime application area concerns a lifts manufacturing industry, targeting condition monitoring at both the actual production as well as at aspects of the end product operation. Section 2 discusses the use of photonics technologies and sensors in condition and process monitoring tasks. Section 3 outlines the adopted e-maintenance architecture, wherein the optical sensing solutions are to be integrated. The photonics sensing platform is presented in Sect. 4, followed by the description of the industrial application testing in Sect. 5. The final section is the conclusion.

2 Photonics Technologies and Sensors in Condition and Processes Monitoring

Photonics technology is essentially an emerging branch of Applied Physics that studies and exploits fundamental physical phenomena of light-matter interaction at micro and nanoscale. It is being recognised as one of the six Key Enabling Technologies (KET) which hold great potential for making a significant impact on industrial innovation leading to smart, sustainable and inclusive growth in Europe, with the other being nanotechnology, micro and nanoelectronics, advanced

materials, industrial biotechnology and advanced manufacturing systems [8]. Photonics may serve an extremely broad range of applications, such as high-speed telecommunication networks, sensing, defence, medicine and energy fusion, which are anticipated see a dramatic penetration into everyday life applications. Photonics technology is becoming quite diverse in terms of employed photonic platforms and materials. Thus, although it is beyond the scope of the current paper to offer a comprehensive classification without including technical content, it can nonetheless be argued that in the area of photonics sensing with industrial applicability, the main currently involved technologies are rather confined in compact waveguiding photonic platforms like optical fibers and integrated optics [9]. Development of new advanced materials together with novel design strategies [10] have enabled the development of highly functional optical sensors that offer versatile operation characterized by unique sensitivity, compactness, reliability, electromagnetic immunity and relatively low cost. Furthermore this photonics platform exhibits the additional inherent capability of fiber-optic based transmission allowing high speed interconnection of multiple remote sensors in a single management centre, combining uniquely the transmission medium and the sensor heads, thereby leading to more efficient monitoring schemes [11].

Despite the huge potential of photonics and fiber optic sensors, their deployment in real industrial environments and infrastructure for asset and processes monitoring is still relatively limited. Photonics sensors technology is still relatively new and the lack of high-level standardization impedes their anticipated wider deployment. A number of efforts in Institute of Electrical and Electronics Engineers (IEEE) and International Electrotechnical Commission (IEC) are currently under way for the standardization of fiber optic and in general photonic sensors.

A number of different physical principles are currently employed in fiber and guided photonic sensors for measuring a variety of parameters, leading to technologies like interferometric sensors, resonators, plasmon sensors, Mach Zehnder interferometers, specially designed photonic crystal fibers, Bragg and long-period gratings. The majority however of photonic sensors lie on the special category of Bragg Grating-based sensors due to the fact that those sensors can be adapted and implemented so that they can be used in a number of different applications, ranging from infrastructures' Structural Health Monitoring (SHM) [9] to monitoring of chemical processes in chemical and pharmaceutical industry [12]. Photonic sensors can find a broad range of applications on various monitoring tasks, as briefly discussed next.

Industrial Applications. The complete electrically passive nature of photonic sensors gives them an added value of extreme significance in demanding industrial applications, wherein monitoring and measurements at harsh environments, such as high-temperature environments inside the oil of power transformers, is required and traditional electric sensors fail to operate safely. Knowledge of the local temperature distribution present in high-voltage, high-power equipment, such as generators and transformers, is essential in understanding their operation and in verifying new or modified products. Relevant applications can be found for leakage detection in pipelines, fault diagnostics and detection of magnetic/electrical field anomalies in

power distribution systems and intrusion alarm systems, as well as in monitoring processes such as composite curing, injection molding and extrusion, or oil drilling. All these are important applications areas for fiber technology. In food and pharmaceutical industry there is an increasing trend for continuous real time monitoring of manufacturing processes and quality control. Typically, in this sector, there is no real time monitoring but sampling and post-control of the products leading to reduced efficiency time and cost-wise. Accurate and low cost photonic approaches enabling the identification of predetermined concentrations at critical points of the processes, could benefit this areas, as for example seen in [12] where close monitoring of the fermentation process in food industry was successfully applied.

Physical Sensing and Structural Health Monitoring. Fiber optic sensors for SHM can offer a range of physical measurements, such as strain, load, rotation, vibration, displacement, force, load, torque, acceleration, pressure and flow, with applicability in structures, bridges, tunnels, pipelines, storage tanks, ships, dams, highways, aircraft wings, spacecraft fuel tanks and oil platforms. Measurement also of load and displacement changes in underground excavations of mines and tunnels is vital for safety monitoring. Many times for the monitoring of a single system a huge variety of sensing parameters are required to provide the full condition. For example, in aircraft propulsion system monitoring, information is required for linear position, temperature, fuel flow, gas pressure, hydraulic pressure, fuel pressure, rotary speed, flame detection, vibration, fluid level etc. Optical sensors could provide the advantage of different sensors implementation in a unified photonics platform with obvious advantages in monitoring systems implementation.

Chemical and Biochemical Sensing. Optical sensor architectures combined with highly selective polymers and nanostructured materials as sensing layers are also suitable for organic compounds' sensing (pollutants, agrochemicals and nerve agents, explosives, drugs and pharmaceuticals and miscellaneous organics), as well as for metals and ion inorganic sensing. Combining photonics sensors and microfluidic cells have demonstrated sensing chips for accurate detection of bacteria, viruses, and toxins. The sensing could be performed as label free or by using biological components in the recognition process, such as enzymes, antibodies, and oligonucleotides.

3 The WelCOM e-Maintenance Architecture

The main ICT enablers for e-maintenance are web-based and semantic maintenance [13], taking into account interoperability considerations (typically the Machinery Information Open System Alliance, MIMOSA, standard—www.mimosa.org), mobile and context-aware computing [2], cloud computing and smart data management [14], as well as wireless sensing and identification [3, 4]. An e-maintenance architecture has been developed that seeks to employ such technologies to vertically integrate data and processes from the shop floor up to the level of maintenance management [7]. To do so the architecture comprises several components as shown in Fig. 1 [14].

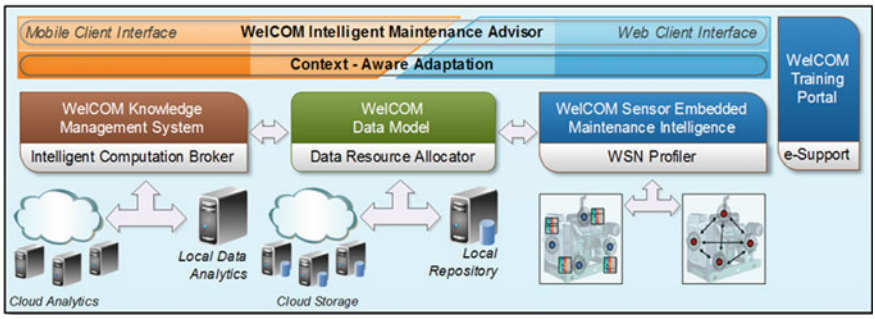


Fig. 1 The WelCOM e-maintenance architecture

At the lower level the architecture's main functional block is that of a smart node in a wireless sensor network infrastructure, constituting the sensor embedded maintenance intelligence of the WelCOM (Wireless Sensor Networks of Optimal Lifecycle Asset Management project, welcom-project.ceti.gr) architecture. A common data model, conforming to the MIMOSA schema with some extensions is unifying data exchange between lower layer and higher layer components, such as the WelCOM intelligent maintenance advisor. This component exploits data and knowledge to export services via context-adaptive interfaces to web or mobile clients. Contextualised-support is offered to users via an e-support and training component. In this setting the concept of context is driving adaptation, that is supporting the delivery of relevant data and services to the apparent context of each service request (Fig. 2). In mobile asset and maintenance management, context can fall under different categories, namely user, social, environment, system and service

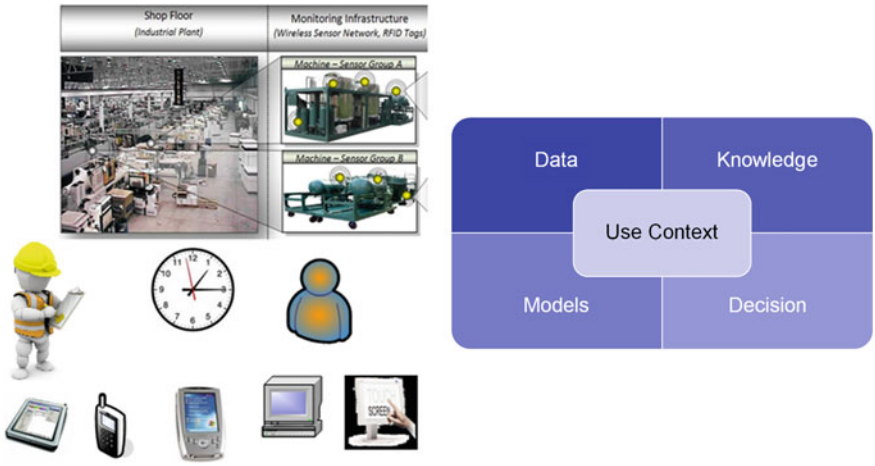


Fig. 2 The concept of context as a driver of adaptation in asset management services

context [14]. A key underlying concept is the decoupling of the computational infrastructure for detection, diagnosis, prognosis and maintenance support. Event detection is thus delegated to the level of a smart node or gateway, while higher level condition monitoring and maintenance management processes are handled at higher levels in the maintenance advisor (Fig. 1). This architecture is powered at the lower level by smart wireless components performing the measurement tasks but also the basic level of data processing, that of pre-processing and event detection. In typical wireless sensing solutions this is hampered by the limited resources of the sensing node, primarily related to energy consumption. Sensors and sensor nodes consume power for the sensing, computing and transmission Radio Frequency (RF) operations, with the latter typically being the source of most power consumption.

A key innovative element of the WelCOM e-maintenance architecture is the introduction of optical sensing elements, based on photonic technologies, at the lower sensing end, a choice that significantly increases power autonomy and versatility of deployment. The approach is described in the next paragraph.

4 The WelCOM Photonic Sensing Platform

The WelCOM architecture incorporates solutions for asset monitoring through a distributed network of sensors that are wirelessly interconnected and managed. The innovation challenge in terms of sensing technology is the development and employment of a suitable photonics platform that could be efficiently embedded towards the implementation of hybrid wireless sensing nodes with actual photonic sensing elements. For the development of those hybrid Wireless Optical Sensors (WOS) a careful consideration of the required specifications is critical. More specifically two important and generic requirements had to be fulfilled for the successful adoption of a photonics sensor in a wireless sensing node:

- The power consumption of the driving circuits of the light transmitter and receiver must be low, in order to have rather big intervals of time between the battery changes, or even autonomous operation by other energy harvesting schemes.
- The interrogation system of the photonic sensor must be compact, simple and inexpensive.

As mentioned in Sect. 2, the major representative of photonics and optical fiber sensors is the class of Bragg Gratings-based sensors. Despite the relatively low implementation cost of Bragg Gratings (BG) based sensors, their use is associated with a relatively high cost as the operation of BG sensors require expensive diode laser sources, optical circulators, isolators, and most importantly expensive wavelength interrogators as the response and the measurement of those sensors take place at the spectral domain. Despite the tendency lately for miniaturization of

wavelength interrogators and the reduction of their prices their cost remains prohibitively high (above 5 K€) for a single sensor. This characteristic makes the current class of BG-based sensors unsuitable for the development of autonomous wireless enabled optical sensors. Furthermore, the only feasible scheme for interrogation of wireless enabled sensors would be the absolute amplitude interrogation scheme.

A potentially viable solution to aid the incorporation of photonics technology sensors in wireless systems has been proven [15–17] to be the low cost Polymer/Plastic Optical Fibers (POF), that combine the required key features of low cost, low power consumption, small size, light weight, immunity to electromagnetic interference (EMI), low interrogation complexity, environmental ruggedness, as well as great sensing operational range and versatility. In contrast to conventional widely used silica (glass) fibers, the plastic optical ones are much easier to handle and more robust towards breakage, due to their higher elasticity. This characteristic was also critically important for POFs use in structural health monitoring applications, as the silica fiber Bragg Grating (FBG) sensors exhibit a rather high failure ratio -due to breakage incidents during installation- of more than 30 %.

Polymer or Plastic Optic Fibers (POFs) is a type of fiber that is mainly used for short-length communication applications. Since their first appearance, POFs show much higher power attenuation compared to their silica counterparts, but with the development of new fabrication methods and polymer materials, the attenuation has been greatly reduced. Briefly a PMMA (Poly(methyl methacrylate)) fiber had an attenuation of 500 dB/km on the late sixties, when was firstly introduced, and now the attenuation is around 150 dB/km. Different types of POFs have also been introduced, like perfluorinated polymers (CYTOP™) that present much lower attenuation characteristics. Their inherent advantages, compared to silica fibers, are the much larger diameter (0.25–3 mm), the reduced cost of the fiber and the lower cost of the tranceivers and optical components. They are also much easier to handle, terminate, connect and are more flexible. They also exhibit excellent chemical and weather resistance and do not easily corrode when in direct contact with a liquid medium. As mentioned above for sensing applications, due to their large core, POFs can be easily manipulated. In many cases, at the sensing region, the jacket of the fiber is completely removed and usually either the fiber is side-polished, so the cladding is removed leaving the core in direct contact with the environment or even a complete segment of the core is removed. The typical POFs used for the need of the WelCOM project are highly multimode fibers of 1 mm diameter with an additional protective jacket of total diameter 2.2 mm (Fig. 3). The typical transceiver unit that is a Fiber Optic Driver Circuit Board (FODCB), utilizes a low cost LED of 200 μW at 650 nm, a photodetector, an amplifier for amplifying the measured voltage and the appropriate electronic driving circuits. The output of the amplifier is connected to the A/D converter of the wireless sensor node for further processing. The FODCB is supplied by 2 AA batteries of 3 V in total and can be seen in Fig. 4.

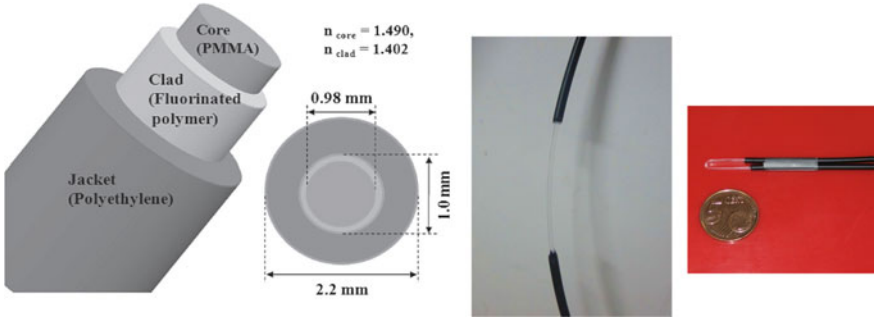


Fig. 3 Schematic of a jacketed POF. POF with stripped jacket and exposed core. Functional U-Bent POF as sensor head

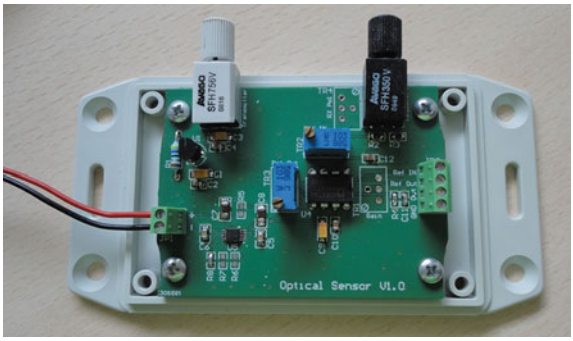


Fig. 4 Prototype of a fiber optic driving circuit board (FODCB)

5 Pilot Testing

The proposed photonic platform of POFs is essentially a very simple guiding medium but with a very high potential for customization, enabling rapid prototyping techniques for implementing customized and highly specialized sensors. Through the WelCOM project the adaptability of the POF platform in various sensing and monitoring requirements, from strain and displacement measurements to monitoring of chemical properties and quality of coolant and lubricant fluids used in machinery, such as in honing machines for lift equipment cylinder finishing (Fig. 5). An important advantageous characteristic of POFs is their high elasticity or their low Young modulus that allows monitoring of relatively high strains in physical structures, in contrast to silica fibers that are restricted to 0.5 % strain limits. The strain monitoring capability of POFs was thoroughly studied, observing the strain in various applications. POFs were embedded in stretchable elastomeric materials in order to monitor dynamically their applied stress and strain. POF-based strain sensors were devised either by longitudinally embedding the POF into the samples or by incorporating alternatively the POF together with a deformable closed-circular loop [18, 19]. The applied strain resulted in elongation or

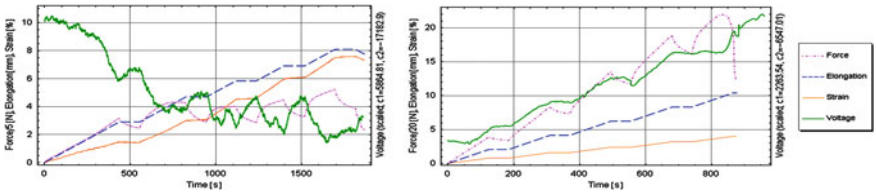


Fig. 5 Testing at Klemmann Lifts. *Left* The lift testing facility. *Right* Metal finishing machine

deformation of embedded fibers leading to corresponding transmitted light loss, which offers an indication of the applied strain. Figure 6 presents schematically the dynamic behaviour of POFs in those two different tensile tests and architectures, demonstrating distinct POF responses that can match different monitoring requirements. Similar techniques can be applied in specific cases of monitoring strain in elevators' wire ropes and relevant applications. Further to their intrinsic characteristics, such as elasticity, the functionality of POFs could be further enhanced as chemical sensors by the use of novel nanostructured sensitive optical materials as overlayers on top of their cladding or core [20, 21]. Selection of proper materials can produce highly customizable POF sensors for a variety of applications in chemical and biochemical detection covering industrial needs of process monitoring in food or pharmaceutical industry. Modern laser micromachining techniques [21] can accurately process POFs surface forming suitable features that can act as hosting cavities for efficient functionalisation [22, 23] of POFs' with sensitive materials. Figure 7 shows the development of POF sensor heads with sensitive overlayers and the development of a laser-induced sensing window on a POF. Based on the chemical sensing capabilities of the POF platform and considering the industrial monitoring requirements of Kleemann Lifts S.A, an efficient sensor for

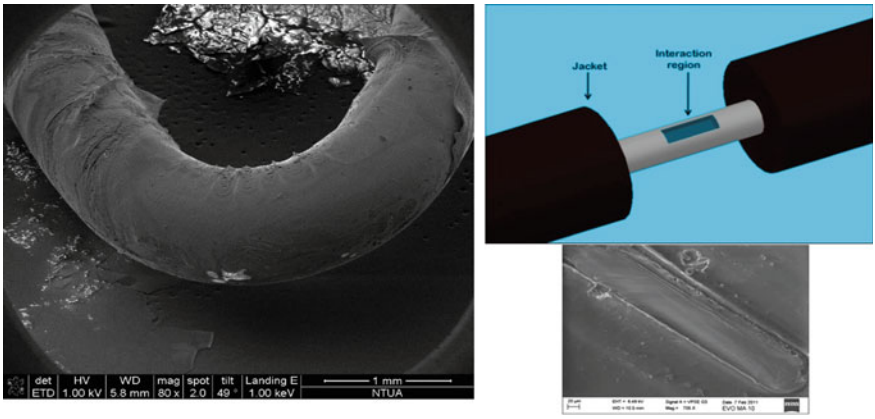


Fig. 6 Response of a POF (*green continuous line*) under a tensile test. *Left* Dynamic behaviour of POF, under a strain ramp, due to its intrinsic elasticity. *Right* Dynamic behaviour of a deformable closed loop POF due to resulted bend loss (see Ref [18])

Fig. 7 *Right* A scanning electron microscope-SEM picture of a U-Bent POF sensing head deposited with sensitive material. *Left* A schematic of a laser micromachined sensing window on the surface of a POF. SEM picture of an actual sensing window suitable for hosting sensitive material

monitoring the aging and the quality of coolant fluids in metal cylinders' finishing machinery (e.g. Honen, Bossi) has been developed. The aging and properties of the coolant was related to the development of biological content/population and the associated change of pH, enabling thus a POF-based sensor for monitoring the coolant's properties.

6 Conclusion

The potential of involving photonics sensor technologies in wireless condition monitoring has been proposed in this paper. Whereas mobile/wireless computing and sensing have increased the flexibility of condition monitoring to integrate and serve mobile users, as well to produce easily customisable and deployable monitoring solutions, power efficiency has remained an issue of concern in wireless monitoring. A promising alternative is the incorporation of photonics technologies, which are applicable to diverse application needs, due to their versatility, low involved costs, installation and operational flexibility, as well as unique safety and reliable operation characteristics in real industrial environments of excessive electromagnetic interference and noise. This paper presented current work on an architecture for condition monitoring, which exploits the added value of inherently EMC-safe optical technology. Promising results are obtained by applying optical sensing to both production equipment as well to the end product of a lifts manufacturing industry. The photonic platform of plastic optical fibers was selected as the most versatile platform suitable for rapid customization and prototyping, able to serve quite diverse sensing and monitoring applications, ranging from physical parameters such as strain and displacement in machinery parts to chemical/biochemical monitoring of industrial-grade coolants. Use of novel nanostructured optical materials together with laser-based micromachining techniques enabled the functional enhancement through rapid prototyping of the optical fibers towards highly customizable sensors.

Acknowledgements The collaboration with all project partners and especially with Kleemann Lifts for providing the application case, as well as Atlantis Engineering and Prisma Electronics for contributing to the application scenario setup and integration, within the context of the GSRT project 09SYN-71-856, 'WelCOM', is gratefully acknowledged.

References

1. Emmanouilidis C, Komonen K (2013) Physical asset management practices in industry: comparisons between Greece and other EU countries. Advances in production management systems. Sustainable production and service supply chains, IFIP advances in information and communication technology, vol 415. Springer, New York pp 509–516
2. Emmanouilidis C, Liyanage JP, Jantunen E (2009) Mobile solutions for engineering asset and maintenance management. J Qual Maint Eng 15(1):92–105 (Emerald)
3. Liyanage JP, Lee J, Emmanouilidis C, Jun N (2009) Integrated e-Maintenance and Intelligent maintenance systems. In: Ben-Daya M, Duffuaa SO, Raouf A, Knezevic J, Ait-Kadi D (Eds) Handbook of maintenance management and engineering. Springer, New York, pp 499–544
4. Emmanouilidis C, Pistofidis P (2010) Wireless Condition Monitoring and Embedded Novelty Detection. In: Amadi-Echendu JE, Brown K, Mathew J, Willet R, Mathew J (eds) Definitions, concepts and scope of asset management. Eng Asset Manage Rev, vol 1. Springer, NEw York, pp 195–238
5. García YR, Corres JM, Goicoechea J (2010) Vibration detection using optical fiber sensors. J Sens 2010(Article ID 936487):12. (Hindawi) doi:10.1155/2010/936487
6. Zhu Y-K, Tian G-Y, Lu R-S, Zhang H (2011) A review of optical NDT technologies. Sensors 11:7773–7798 (MDPI)
7. Pistofidis P, Emmanouilidis C, Koulamas C, Karampatzakis D, Papathanassiou N (2012) A layered E-maintenance architecture powered by smart wireless monitoring components. In: Proceedings of the 2012 IEEE international conference on industrial technology (ICIT 2012). Athens, Greece 19-21/3, IEEE, pp 390–395
13. Kiritsis D (2013) Semantic technologies for engineering asset life cycle management. Int J Prod Res 51:7345-7371 (Taylor & Francis). doi: 10.1080/00207543.2012.761364
12. Pistofidis P, Emmanouilidis C (2013) Profiling context awareness in mobile and cloud based engineering asset management. Advances in production management systems. Competitive manufacturing for innovative products and services. IFIP AICT, vol 398. Springer, New York, pp 17–24
8. Key Enabling Technologies Final Report (2011) High-Level Expert Group on key enabling technologies. European Commission
9. Pruneri V, Riziotis C, Smith PGR, Vasilakos A (2009) Fiber and integrated waveguide-based optical sensors. J Sens 2009(Article ID 171748)
10. Riziotis C, Vasilakos A (2007) Computational intelligence in photonics technology and optical networks: a survey and future perspectives. Inf Sci J 177:5292–5315 (Elsevier)
11. ZuDe Z, Quan L, QingSong A, Cheng X (2011) Intelligent monitoring and diagnosis for modern mechanical equipment based on the integration of embedded technology and FBGs technology. Measurement 44:1499–1511
14. Sparrow IJG, Smith PGR, Emmerson GD, Watts SP, Riziotis C (2009) Planar Bragg Grating sensors—fabrication and applications: a review. J Sens 2009:Article ID 607647
15. Riziotis C, Dimas D, Katsikas S, Boucouvalas AC (2010). Photonic sensors for autonomous wireless sensing nodes. In: Proceedings of the 23rd international congress on condition monitoring and diagnostic engineering management. COMADEM 2010. 28/6-2/7. Nara, Japan, Sunrise Publishing Limited, pp 669–676
16. Dimas D, Katsikas S, Boucouvalas AC, Riziotis C (2011). Low cost, autonomous and wireless enabled liquid level sensor based on a multi-segmented polymer optical fiber. SENSOR+TEST

conferences 2011, OPTO 2011, 7–9 June 2011, Nurnberg Exhibition Centre, Germany, pp 145–150
17. Dimas D, Katsikas S, Boucouvalas AC, Riziotis C (2011). Wireless-enabled photonic sensor for liquid level and distributed flood monitoring. In: Proceedings of the 24th international congress on condition monitoring and diagnostic engineering management. COMADEM 2011. 30/5-1/6. Stavanger, Norway, pp 434–444
18. Riziotis C, Eineder L, Bancallari L, Tussiwand G (2013) Fiber optic architectures for strain monitoring of solid rocket motors' propellant. Sens Lett 11(8):1403–1407
19. Riziotis C, Eineder L, Bancallari L, Tussiwand G (2013) Structural health monitoring of solid rocket motors' propellant using polymer optical fibers. In: Proceedings of the 2nd international conference on materials and applications for sensors and transducers IC-MAST 2012, Budapest, Hungary, 24–28 May 2012. Key Engineering Materials, vol 453. pp 360–363
20. Athanasekos L, Pispas S Riziotis C (2012) Novel block copolymers for multi-agent detection using polymer optical fiber. In: SPIE photonics Europe, 16–19 April 2012, Square Brussels Meeting Centre, Brussels, Belgium, Proceedings of SPIE 8426, 842615
21. Athanasekos L, Aspiotis N, El Sachat A, Pispas S, Riziotis C (2013) Novel approach for lysozyme detection employing block copolymer overlayers on plastic optical fibers. In: Proceedings of the 2nd international conference on materials and applications for sensors and transducers IC-MAST, Budapest, Hungary, 24–28 May 2012. Key Engineering Materials, vol 543, pp 385–388
22. Athanasekos L, Dimas D, Katsikas S, Pispas S, Vainos N, Boucouvalas AC, Riziotis C (2013) Laser microstructuring of polymer optical fibres for enhanced and autonomous sensor architectures. Proc Eng 25:1593–1596
23. Athanasekos L, El Sachat A, Pispas S, Riziotis C (2014) Amphiphilic diblock copolymer based multi-agent photonic sensing scheme, Journal of Polymer Science Part B: Polymer Physics, 52:46-54

A Combined Life Prediction Method for Product Based on IOWA Operator

Lei Feng, Xiaoyang Li, Tongmin Jiang and Xiangjun Dang

Abstract As the rapid development of modern technology, industrial companies have to manufacture high-reliable and long-lifetime products. How to evaluate these indexes is an urgent problem to be solved. Utilizing the product degradation information may be an effective way to solve this issue. However, most of lifetime prediction models in use are mainly based on single prediction model with the shortage of low robustness and accuracy. In this chapter, a combined prediction method based on the performance degradation data by using the induced ordered weighted averaging operator (IOWA) is proposed. We select two better prediction models, which are time series model and BP neural network, to predict the degradation path of product respectively. The IOWA operator can build a new combination prediction method which can overcome the defect of fixed weight coefficients of the traditional combined method. This method can update the weight coefficients dynamically according to the prediction precision of each model. Then the objective function of the error square sum is established with weight coefficients used to combine these prediction methods and integrate the prediction results.

Keywords Combined prediction · Time series model · BP neural network · IOWA operator · Performance degradation data

L. Feng · T. Jiang · X. Dang
Products and Environmental Engineering Research Center, Beihang University, 37 Xueyuan Road, Haidian District, 100191 Beijing, China
e-mail: fenglei3714@hotmail.com

T. Jiang
e-mail: jtm@buaa.edu.cn

X. Dang
e-mail: xiangjun@dse.buaa.edu.cn

X. Li (✉)
Science and Technology on Reliability and Environmental Engineering Laboratory, Beihang University, 37 Xueyuan Road, Haidian District, 100191 Beijing, China
e-mail: leexy@buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_119

1 Introduction

With the development of science and technology, the long-life & high-reliability products become more common in the field of astronautics and aeronautics. Thus how to evaluate the lifetime and reliability of those products is an urgent problem to be solved. The accelerated degradation testing (ADT) attempt to obtain the degradation information under the condition of limited time and funds by using more severe stress than the normal level. Therefore, to utilize the product degradation information from ADT may be an effective way to resolve this issue. However, most of lifetime prediction models in use are mainly based on single prediction model with the disadvantage of low robustness and accuracy, so it cannot get the high credible lifetime and reliability of the product. Then this inaccurate result may lead to serious consequence. But a new prediction model combining several single models may overcome this defect.

Combined forecast proposed by Bates and Granger in 1969 is a prediction method with the thought of combining more than one single model, which can reduce and disperse the uncertainty of the single prediction model so as to improve the accuracy of the prediction [1]. Then combined forecast has aroused great attention in various field. Xie et al. have studied lifetime prediction of the electronic devices for the vehicles, and put forward a combined prediction model based on exponential regression model and grey theory [7]. Reng et al. have applied combined prediction method to predict the residual life of aero-engine based on the performance degradation [4]. Tang et al. minimize the objective function of the prediction error square sum to determine the weighting coefficient of the optimal combination prediction model [5]. Long et al. combine the induced ordered weighted averaging operator (IOWA) with Markov chain (MC) to construct an IOWA-MC forecasting model to remedy the defects of traditional method for forecasting the long term power load [3]. Xie et al. also apply the combined prediction model based on the improved IOWA to predict the maintenance cost of the ship equipment [8].

In this chapter, the author selects two better prediction models, which are time series model and BP neural network, to predict the degradation path of the product separately, then IOWA operator is used to build a new combination model which can overcome the defect of fixed weight coefficients of the traditional combined method, this method can update the weight coefficients dynamically according to the prediction precision of each model. Then the objective function of the error square sum is established with weight coefficients which are used to combine these prediction methods and integrate the prediction results. And the objective function is minimized with the genetic algorithm, which has the characteristics of fast global optimization and uses to determine the weight coefficients for each of the prediction methods.

2 Single Prediction Models

This part will introduce two single life prediction models used to build the combined prediction model, which are time series prediction model and BP neutral network model. Then the detailed modelling procedure is as follows.

2.1 Time Series Prediction Model

As the product influenced by random factors in practical engineering, the performance degradation path is also a stochastic sequence. Also the degradation path usually presents a monotonic degradation trend determined by the physical and chemical mechanism of the product, which is usually expressed by multiplying the degradation rate and linear or nonlinear logarithmic function, exponential function, power function and so on [6]. Set

$$Y_t = F_t + C_t + R_t,\ t = 1, 2, \cdots \tag{1}$$

Here $F_t$is the trend part, which is monotone function determined by degradation mechanism of the product. C_t is the periodic part determined by the internal and external factors from the product. R_t is the zero-mean stochastic part. Among which, the trend part F_t can be expressed as follows:

$$F_t = bg(t) + y_0 \tag{2}$$

Here, b is the degradation rate of the degradation path, $g(t)$ is the monotone function, y_0 is the initial value of the performance degradation.

The periodic part is usually several cosine signal, and the periodic part can be expressed like this:

$$C_t = \sum_{j=1}^{k} a_j \cos(w_j t + \varphi_j) \tag{3}$$

Here, $k$is the number of angular frequency, a_j is the amplitude of the jth angular frequency, $w_j$is the jth angular frequency, φ_j is the jth phase angle.

After the degradation path extracts the trend part and periodic part, the remaining stochastic part with linear and stationary characteristic can be modelled with the stationary time sequence model. The stochastic data test through stationarity test does zero mean processing, then it can be expressed by autoregressive model as follows:

$$R_t = \sum_{j=1}^{p} \eta_j R_{t-j} + \varepsilon_t$$
$$E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, Cov(\varepsilon_t, \varepsilon_{t-i}) = 0, \forall i \geq 1 \quad (4)$$

Here, p is the autoregressive model order, η_j is the autoregressive coefficient, ε_t is the white noise following normal distribution $N[0, \sigma_\varepsilon^2]$.

2.2 BP Neutral Network Model

BP neutral network is a multilayer perception based on BP algorithm, which is composed of the input layer, the middle layer and the output layer. The middle is also called the hidden layer with one or more layers. Figure 1 is the BP neural model with three layers [2].

Among which $(x_1, x_2, \cdots, x_m)$ is n dimensional input vector, z is one dimensional output vector, w_{ik} is the connection weight between the input layer and the middle layer, v_{jt} is the connection weight between the middle layer and the output layer.

3 A Combination Forecasting Model Based on IOWA Operator

The combined prediction method based on IOWA is introduced. Firstly, the IOWA operator can build a new combination forecast method which can overcome the defect of fixed weight coefficients of the traditional combined method, this method can update the weight coefficients dynamically according to the prediction precision of each model. Then the objective function of the sum of the error squares is established with weight coefficients which are used to combine these prediction methods and integrate the prediction results.

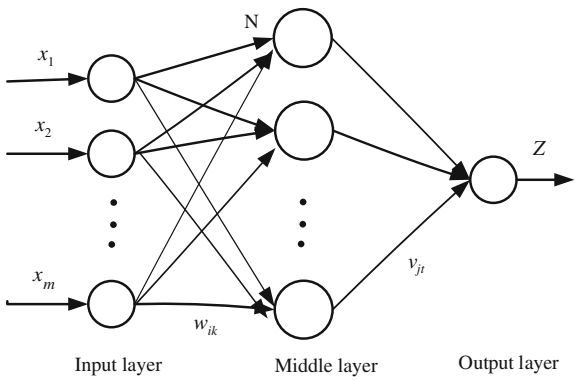


Fig. 1 Neural network structure diagram

3.1 The Introduction of IOWA Operator

Assuming $(a_1, x_1), (a_2, x_2), \cdots, (a_m, x_m)$ are two-dimensional arrays, set:

$$IOWA_L(\langle a_1, x_1\rangle, \langle a_2, x_2\rangle, \cdots, \langle a_m, x_m\rangle) = \sum_{i=1}^{m} l_i x_{a-index(i)} \tag{5}$$

Among which, $L = (l_1, l_2, \cdots, l_m)^T$ is the weighted vector of IOWA and meets $\sum_{i=1}^{m} l_i = 1, l_i \geq 0, i = 1, 2, \cdots, m$. The function $IOWA_L$ is a m-dimensional induced ordered weighted averaging operator caused by $a_1, a_2, \cdots, a_m$, denoted by IOWA operator, and a_i is called induced value of x_i, $a - index(i)$ is the subscript of the ith largest number among $a_1, a_2, \cdots, a_m$,which are listed in a descending order.

3.2 The Combination Forecast Based on IOWA Operator

Suppose $\{x_t, t = 1, 2, \cdots, N\}$ is the observed sequence of degradation information, and m available single prediction models predict the degradation value for the future. Here x_{it} is prediction value of the ith prediction method at moment t, $i = 1, 2, \cdots, m, t = 1, 2, \cdots, N.L = (l_1, l_2, \cdots, l_m)^T$ is the weighted vector of IOWA and meets $\sum_{i=1}^{m} l_i = 1, l_i \geq 0, i = 1, 2, \cdots, m$.

Set

$$a_{it} = \begin{cases} 1 - |(x_t - x_{it})/x_t|, & when|(x_t - x_u)/x_t| < 1 \\ 0, & when|(x_t - x_u)/x_t| \geq 1, \end{cases} \quad i = 1, 2, \cdots, m,\ t = 1, 2, \cdots, N. \tag{6}$$

Here, a_{it} shows the fitting accuracy of the ith prediction method, obviously $a \in [0, 1]$. If the fitting accuracy a_{it} is treated as induced value of the prediction value x_{it}, thus the fitting accuracy of m single prediction methods at moment t and their corresponding forecasting values constitute m two-dimensional arrays $(\langle a_1, x_1\rangle, \langle a_2, x_2\rangle, \cdots, \langle a_m, x_m\rangle)$ [9].

Set

$$IOWA_L(\langle a_{1t}, x_{1t}\rangle, \langle a_{2t}, x_{2t}\rangle, \cdots, \langle a_{mt}, x_{mt}\rangle) = \sum_{i=1}^{m} l_i x_{a-index(it)} \tag{7}$$

$IOWA_L$ indicates the combined prediction value based on IOWA, which is generated by the fitting accuracy sequence $a_{1t}, a_{2t}, \cdots, a_{mt}$ at moment t. Apparently, the most characteristic of combined forecasting based on IOWA lies in that

weighting coefficients are independent of single prediction methods, but they are closely related to the fitting accuracy of each single prediction method at different moments, the higher the fitting accuracy, the greater the weighting coefficients.

Let $e_{a-index(it)} = x_t - x_{a-index(it)}$, then the error square sum of combined forecasting error is:

$$S = \sum_{t=1}^{N} (x_t - \sum_{i=1}^{m} l_i x_{a-index(it)})^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} l_i l_j (\sum_{t=1}^{N} e_{a-index(it)} e_{a-index(jt)}) \quad (8)$$

In terms of the criteria of minimizing the error square sum, the combined model based on the IOWA operator can be expressed as follows:

$$\begin{aligned} &\min S(L) = \sum_{i=1}^{m} \sum_{j=1}^{m} l_i l_j (\sum_{t=1}^{N} e_{a-index(it)} e_{a-index(jt)}) \\ &s.t. \begin{cases} \sum_{i=1}^{m} l_i = 1 \\ l_i \geq 0,\ i = 1, 2, \ldots, m. \end{cases} \end{aligned} \quad (9)$$

After the optimal IOWA weighting coefficients of combined forecasting are obtained, namely $L^* = (l_1^*, l_2^*, \cdots, l_m^*)^T$, according to the principle of continuity, it can be utilized to predict in forecasting intervals $[N+1, N+2, \cdots]$, the equation is as follows.

$$\begin{aligned} &IOWA_{L^*}(\langle a_{1t}, x_{1t}\rangle, \langle a_{2t}, x_{2t}\rangle, \cdots, \langle a_{mt}, x_{mt}\rangle) = \sum_{i=1}^{m} l_i^* x_{a-index(it)}, \\ &t = N+1, N+2, \cdots \end{aligned} \quad (10)$$

The value of prediction accuracy sequence $a_{1t}, a_{2t}, \cdots, a_{mt}$in combined forecasting intervals $[N+1, N+2, \cdots]$ is in accordance with recent average fitting accuracy of each single prediction method. To predict the step k of the future, the $N+k$-phase prediction accuracy of the ith prediction method approximately equals to the average fitting accuracy in recent k-phase.

4 Engineering Application

Accelerated degradation data of an SLD at 55 ℃ are utilized to verify the combined prediction method based on the IOWA operator. Now we got the degradation data of the SLD by accelerated degradation testing within 480 h, the raw data shows in Fig. 2.

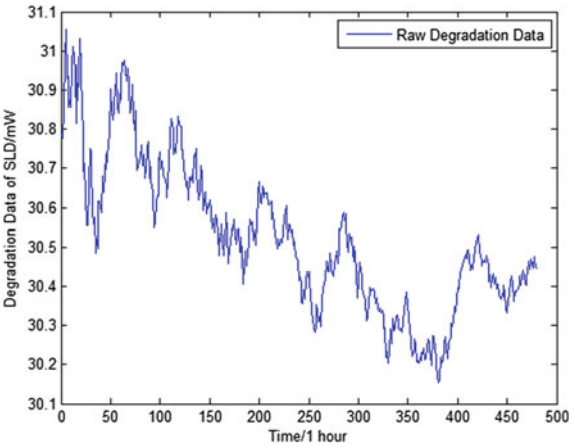


Fig. 2 Raw degradation data of the SLD

The raw data is evenly divided into two parts, the first part of 240 h is used to establish degradation model, and the other part will test the goodness of the model. Now the first part data is selected as modelling data for time series model and learning samples for BP network model. The other part will test the prediction accuracy of the single prediction model and the combined model.

The time series model is built as follows,

$$\begin{aligned} y(t) =& 0.07517 \times \cos(0.09818t - 0.3316) + 0.06436 \times \cos(0.1309t - 1.0853) \\ & - 0.001510t + R_t + 30.854 + 0.04805 \times \cos(0.05236t + 1.0853) \\ R_t =& 0.8054R_{t-1} - 0.05525R_{t-2} + \varepsilon_t \qquad \varepsilon_t \in N(0, 0.07^2) \\ R_1 =& -0.1498 \qquad R_2 = -0.1635 \end{aligned} \tag{11}$$

Among which, the blue curve is the raw degradation data, the red curve is the fitting degradation data, and the green curve is the prediction data (Fig. 3).

Then BP neural network model is built to predict the degradation path, and he BP neural network settings list in Table 1.

Figure 4 is the BP neural network prediction model, the blue curve is the raw degradation data, the green curve is the fitting degradation data, and the red cure is the prediction data.

After the BP neural network model and time series model are established respectively, the fitting accuracy at each moment can be calculated by the fitting data of single prediction model. Then the fitting error of the IOWA combined model are calculated by Eq. (9). The combined forecasting model based on IOWA is obtained on the criteria of minimizing the error square sum. i.e.:

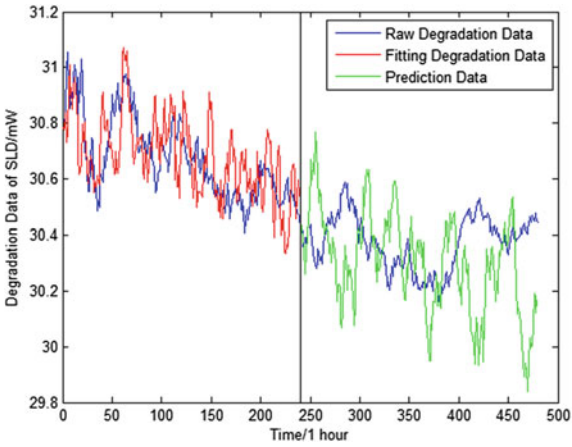


Fig. 3 Time series prediction model

Table 1 BP neural network settings

BP neural network settings	Value or function
Input layer neurons number	1
Middle layer neurons number	6
Output layer neurons number	2
Transfer function	Tansig, tansig, purelin
Training function	Levenberg-Marquardt
Epoch	1,000
Performance	0.001
Validation checks	20

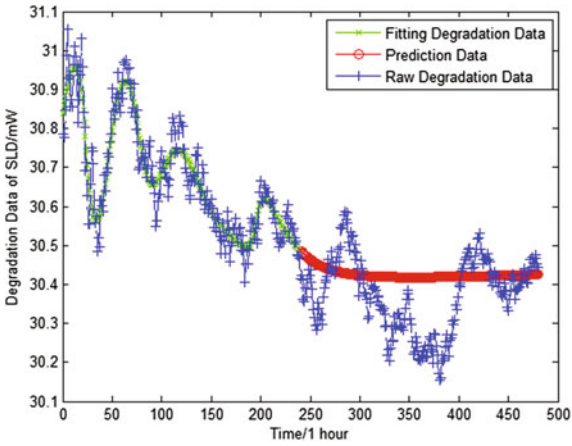


Fig. 4 BP neural network prediction Model

$$\min S(l_1, l_2) = 1.0711l_1^2 + 0.9832l_1l_2 + 4.088l_2^2$$
$$s.t. \begin{cases} l_1 + l_2 = 1 \\ l_1 > 0, l_2 > 0 \end{cases} \quad (12)$$

The objective function is minimized with the genetic algorithm, the optimal weighting coefficients of the model are obtained, $l_1^* = 0.8612$, $l_2^* = 0.1388$. By taking the average fitting accuracy of each single prediction model as the predictive accuracy, the combined forecasting model based on IOWA operator can predict the degradation path.

In order to validate the combined prediction model based on IOWA established above, the observed degradation data of SLD is used to compare the fitting and predictive accuracy of the each single prediction model and the combined model with each other. Fitting and prediction curves of the combined model show in Fig. 5, in addition, the contrast of fitting and prediction accuracy is illustrated in Table 2. From the Table 2, it indicates that the fitting and prediction accuracy of the combined model based on IOWA operators is higher than each single prediction model.

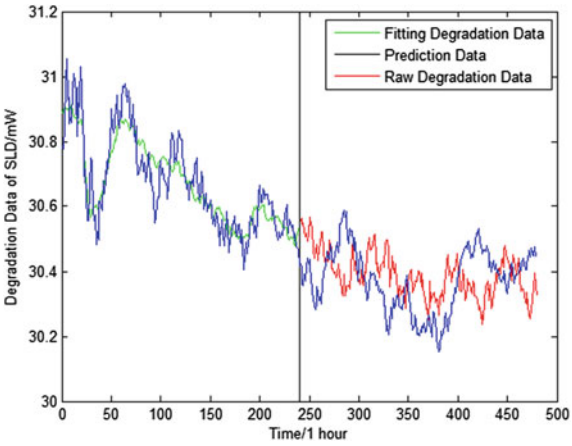


Fig. 5 Combined Prediction Model Based on IOWA operator

Table 2 Comparison of fitting and prediction accuracy of different models

		BP model	Time series model	IO WA combined model
Fitting	Error square sum	1.0711	4.088	1.036
	RMSE[a]	0.0043	0.0084	0.0042
Prediction	Error square sum	2.7164	8.3915	2.4953
	RMSE	0.0069	0.012	0.0065

[a] is short for root mean square error

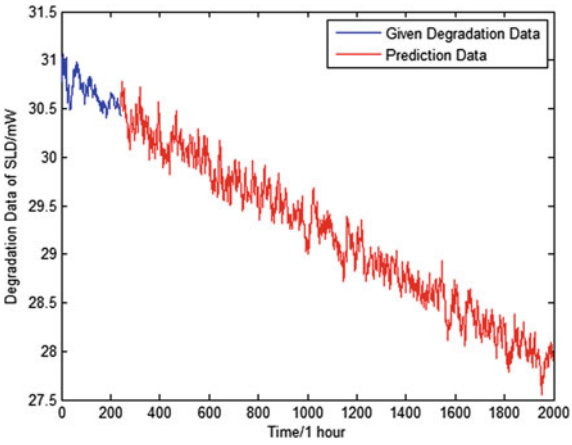


Fig. 6 SLD lifetime prediction by combined prediction model based on IOWA operator

According to the actual failure experience of the SLD, assuming the 93 % initial value to be the failure threshold, the combined prediction model based on the IOWA operator is utilized to predict the degradation path of the SLD, which shows in Fig. 6. Among which, the blue curve is the observed degradation data, and the red curve is the prediction curve. From this figure, the lifetime of the SLD at 55 degrees centigrade can be estimated 2,200 h approximately.

5 Conclusion

In this chapter, the authors establish time series model and BP neural network model as two single prediction models to predict the degradation path of the SLD separately. A new combined prediction model of BP neural network and times series model based on IOWA operator is built, which can overcome the defect of fixed weight coefficients of the traditional method. This method can update the weight coefficients dynamically according to the prediction precision of each model. At last, case study shows that the combined prediction model based on IOWA operator is better than each single prediction model, with higher fitting and prediction accuracy in comparison with two single prediction model.

References

1. Bates JM, Granger CWJ (1969) Combination of forecasts. Oper Res 20(4):451–468
2. Dang XJ, Jiang TM (2013) Degradation prediction based on correction analysis and assembled neural network. J Beijing Univ Aeronaut Astronaut 39(1):42–46
3. Long RH, Ge M (2010) A combination model for medium and long term load forecasting based on induced ordered weighted averaging operator and markov chain. Power Syst Technol 34 (3):150–156

4. Reng SH, Zuo HF (2011) A combined prediction method for the residual-life of civil aviation engines based on performance degradation. Mech Sci Technol Aerosp Eng 30(1):23–29
5. Tang XW (1994) The optimal combination forecasting method and its application. Appl Stat Manage
6. Wang L (2011) Life prediction technology for accelerated degradation testing based on time series analysis. Beijing University of Aeronautics and Astronautics
7. Xie CQ, Zhang YM (2008) Combined forecast of service life about electronic parts of vehicular electronic equipment. J Test Meas Technol 22(3):189–193
8. Xie L, Wei RX (2012) Forecast of ship equipment maintenance cost with improved IOWA combination model. Syst Eng Electron 34(6):1176–1181
9. Yager RR (1998) On ordered weighted averaging aggregation operation in multi-criteria decision making. IEEE Trans Syst Man Cybernet 18(3):183–190

Bayesian Acceptance Sampling Plan for Exponential Distribution Under Type-I and Type-II Censoring

Pengfei Gao, Xiaoyang Li and Xue Song

Abstract This paper studies the design method in engineering for Bayesian acceptance sampling plan under exponential distribution. In reliability theory, the exponential distribution is the most basic and common distribution, plan is widely applicable. First, aiming at the two kinds of truncation method (type-I and type-II censoring), we give the Bayesian posterior expression of the producer and consumer risk according to the Average risk criteria and Posterior risk criteria respectively. Second, in the specific analysis, using the Gamma distribution as the prior distribution of failure rate, we obtain the test plan of different truncated modes and the test plan of different risk criteria as well as the classical methods are compared, the results show that the application of Bayes plan can greatly reduce the test cost. Third, the influence of the prior distribution on the Bayes plan is studied.

Keywords Acceptance sampling · Bayesian method · Producer and consumer risks

1 Introduction

Life test sampling plans (LSPs) are usually used to make a decision on accepting or rejecting batch of products in engineering when the life is an important quality index to measure this product. For some small sample complex products of complex technology, high cost and reliability, First, the classic acceptance plan needs a

P. Gao (✉) · X. Song
Beihang University, Beijing, China
e-mail: wsykty@126.com

X. Song
e-mail: hpusongxue@163.com

X. Li
Science and Technology on Reliability and Environmental Engineering Laboratory, Beihang University, Beijing, China
e-mail: leexy@buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_120

large number of samples, long test time and high experiment cost; Second, a large amount of prior information which appear at every stage of development process have not been effective use. According to the above two kinds of circumstances, Using the Bayes method can make full use of the prior information, save the sample quantity, shorten the test time, thus we can reduce the test cost, improve test efficiency in the premise of ensuring the correct acceptance.

Through the investigation, we know the latest research results about the application of Bayes method in life verification test technology, Balakrishnan [1], etc. design an acceptance sampling plan for the Generalized Birnbaum–Saunders Distribution under time censoring using the Bayes method; Liang and Yang [2] design the Optimal Bayesian acceptance sampling plan for exponential distribution based on hybrid censored samples; Amin [3], etc. design an acceptance sampling plan for the Pareto lifetime model, Lin [4], etc. bring a Bayesian Variable sampling plans for the exponential distribution based on type-I and type-II hybrid censored samples, Fallah Nezhad [5], etc. design a new Bayesian acceptance sampling plan considering inspection errors.

First, aiming at the two kinds of truncation method (type-I and type-II censoring), we give the Bayesian posterior expression of the producer and consumer risk according to the Average risk criteria and Posterior risk criteria respectively. Second, in the specific analysis, using the Gamma distribution as the prior distribution of failure rate, we obtain the test plan of different truncated modes and the test plans of different risk criteria as well as the classical methods are compared. Third, the influence of the prior distribution on the Bayes plan is studied.

2 Both Risk Measure Under Two Types of Risk Guidelines

2.1 Average Risk Criteria

According to the Bayesian Reliability [6], In the discussion of two types of risk calculation method about Average Risk Criteria, based on the principle of probability theory and mathematical statistics, We can get calculation formula of abandon really risk α by derivation as shown as (1):

$$\alpha = P(Z \in D_1 | R \in \theta_0) = \frac{P(Z \in D_1, R \in \theta_0)}{P(R \in \theta_0)} \tag{1}$$

In the formula (1), Z is the decision variables of for life test, D_1 is the rejection region of judging whether the life of the product comply with the requirements or not, R is the life parameters of the product, θ_0 is the value range of life parameters meeting the test requirements.

The calculation formula of in pseudo risk β is shown as formula (2):

$$\beta=P(Z \in D_0|R \in \theta_1) = \frac{P(Z \in D_0, R \in \theta_1)}{P(R \in \theta_1)} \tag{2}$$

In the formula (2), D_0 is the acceptance region of judging whether the life of the product complies with the requirements or not, θ_1 is the value range of life parameters not meeting the test requirements.

2.2 Posterior Risk Criteria

According to definition about Posterior Risk Criteria in the Bayesian Reliability, We can get calculation formula of abandon really risk α as shown as (3):

$$\alpha=P(R \in \theta_0|Z \in D_1) = \int_{\theta_0} p(R|Z \in D_1)dR = \frac{\int_{\theta_0} P(Z \in D_1|R)\pi(R)dR}{\int_{\theta} P(Z \in D_1|R)\pi(R)dR} \tag{3}$$

In the formula (3), θ means universal set of product life parameter R that is $p\ (R|Z\in D_1)$, it means the probability density function of R under the given condition of $Z\in D_1$.

In the Posterior Risk Criteria, the definition of in pseudo risk is show as (4)

$$\beta=P(R \in \theta_1|Z \in D_0) = \int_{\theta_1} p(R|Z \in D_0)dR = \frac{\int_{\theta_1} P(Z \in D_0|R)\pi(R)dR}{\int_{\theta} P(Z \in D_0|R)\pi(R)dR} \tag{4}$$

In the formula (4), $p\ (R|Z\in D_0)$ means probability density function of R on the basis of field test data getting the result of $Z\in D_0$.

2.3 Comparative Analysis of Two Types of Risk

Risk calculation principle of average risk criterion is similar to the classic calculation method of the risks; the biggest difference between them is that the average risk Criteria use priori distribution of the product life parameter, it puts a classic calculation method of the risk as a weighted average based on priori distribution. Different from the average risk criteria, posteriori risk criteria backward different probability between life parameters priori distribution and life level from field tests by means of field test data, and its calculation method is mainly based on a Subjective recognition of a priori distribution of life parameters.

3 Bayesian Acceptance Sampling Plan for Exponential Distribution Under Failure Censoring

3.1 Product Life Indicators Statistical Hypothesis for Exponential Distribution

For exponential distribution products, cumulative distribution function is as follows:

$$F(t) = 1 - \exp(-t/\theta) \tag{5}$$

From the characteristics of exponential distribution we know that its life expectancy and failure rate have the relationship of $\theta = 1/\lambda$, In order to facilitate the selection of prior distribution, this article selects failure rate λ as its life test indicators, Put λ_0 and λ_1 as the test upper and lower of λ, established statistical hypothesis is as follows:

$$H_0: \ \lambda \leq \lambda_0 \qquad\qquad H_1: \ \lambda > \lambda_1 \tag{6}$$

In the formula (6), the value range of failure rate is [0, 1], $\lambda \leq \lambda_0$ means that the life of the product is qualified; $\lambda > \lambda_1$ means that the life of the product is not qualified.

For exponential product test plan under constant truncated, Often written as (r, T), among them, λ is the number of failure, T is the critical time for tests. The decision rule for life test is: Select n products as a sample to test, when the number of samples that a failure happens reaches to r, Stop the test, at this time, if the total test time is greater than T, then accept the original hypothesis, or, accept the alternative hypothesis.

The calculating formula of the total test time t is:

$$t = \sum_{i=1}^{r} t_i + (n - r)t_r \tag{7}$$

In the formula (7), T means the accumulated test time of life validation tests, t_i means the time that the failure I happen; N means the total number of products in the trials in life validation tests, Generally, both the producer and the consumer reach an agreement, so, the main task of failure censoring test design is to choose appropriate r and T.

3.2 The Plan Design Based on Average Risk Criteria

According to the above, the test plan of exponential product has two factors, that is r and T, for failure rate λ of test indicators, according to the Bayes theory, take

the conjugate prior distribution for the Gamma distribution, which is written as $G(a, b)$, that is:

$$\pi(\lambda|a,b) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda} \tag{8}$$

Among, $\Gamma(a)$ is the Gamma function, defined as:

$$\Gamma(a) = \int_0^\infty e^{-x}x^{a-1}\,dx \tag{9}$$

For exponential distribution, its field sampling result is (r, t), R is failure number, T is the total test time of the product r fails, after the failure number r is given, $2\lambda t \sim \chi^2(2r)$, that is:

$$h(2\lambda t) = \begin{cases} \frac{1}{2^r\Gamma(r)} & (2\lambda t)^{r-1}e^{-\lambda t} \quad 2\lambda t > 0 \\ 0 & 2\lambda t \le 0 \end{cases} \tag{10}$$

By the distribution function of a random variable, we can deduce probability density function of t, that is:

$$g(t) = \frac{t^{r-1}\lambda^r e^{-\lambda t}}{\Gamma(r)} \tag{11}$$

Based on average risk criterion and Bayes formula, combined with the prior distribution of the failure rate λ and probability density function of t, we can deduce the calculation formula of two types of risk as follows:

$$\alpha(r,T) = P(t<T|\lambda<\lambda_0) = \frac{\int_0^{\lambda_0}\int_0^T g(t)\pi(\lambda)dtd\lambda}{\int_0^{\lambda_0}\pi(\lambda)d\lambda}. \tag{12}$$

$$\beta(r,T) = P(t\ge T|\lambda>\lambda_1) = \frac{\int_{\lambda_1}^{\infty}\int_T^{\infty} g(t)\pi(\lambda)dtd\lambda}{\int_{\lambda_1}^{\infty}\pi(\lambda)d\lambda} \tag{13}$$

Test plan can be getting by solving the two equations.

3.3 The Plan Design Based on Posteriori Risk Criteria

According to the posteriori risk criteria, and the derivation for prior distribution of exponential products and the probability density function of life t in the previous section, the calculation formula for the first category of risk $\alpha(r, T)$ is:

$$\alpha(r,T) = P(\lambda \le \lambda_0 | t < T) = \int_0^{\lambda_0} p(\lambda | t < T) d\lambda$$
$$= \frac{\int_0^{\lambda_0} P(t < T|\lambda)\pi(\lambda)d\lambda}{\int_0^{\infty} P(t < T|\lambda)\pi(\lambda)d\lambda} = \frac{\int_0^{\lambda_0}\int_0^{T} g(t)\pi(\lambda)dtd\lambda}{\int_0^{\infty}\int_0^{T} g(t)\pi(\lambda)dtd\lambda} \tag{14}$$

The calculation formula for the second category of risk is:

$$\beta(r,T) = P(\lambda > \lambda_1 | t \ge T) = \int_{\lambda_1}^{\infty} p(\lambda | t \ge T) d\lambda$$
$$= \frac{\int_{\lambda_1}^{\infty} P(t \ge T|\lambda)\pi(\lambda)d\lambda}{\int_0^{\infty} P(t \ge T|\lambda)\pi(\lambda)d\lambda} = \frac{\int_{\lambda_1}^{\infty}\int_T^{\infty} g(t)\pi(\lambda)dtd\lambda}{\int_0^{\infty}\int_T^{\infty} g(t)\pi(\lambda)dtd\lambda} \tag{15}$$

Test plan can be getting by solving the two equations.

4 Bayesian Acceptance Sampling Plan for Exponential Distribution Under Time Censoring

4.1 Product Life Indicators Statistical Hypothesis for Exponential Distribution

For exponential distribution products, cumulative distribution function is as follows:

$$F(t) = 1 - \exp(-t/\theta) \tag{16}$$

From the characteristics of exponential distribution we know that its life expectancy and failure rate have the relationship of $\theta = 1/\lambda$, in order to facilitate the selection of prior distribution, this article selects failure rate λ as its life test indicators, put λ_0 And λ_1 as the test upper and lower of λ, established statistical hypothesis is as follows:

$$H_0: \ \lambda \le \lambda_0 \qquad\qquad H_1: \ \lambda > \lambda_1 \tag{17}$$

In the formula (17), the value range of failure rate is [0, 1], $\lambda \le \lambda_0$ means that the life of the product is qualified, $\lambda > \lambda_1$ means that the life of the product is not qualified.

For exponential product test plan under time censoring test, often written as (c, T), among them, c is the number of failure, T is the critical time for tests. The decision rule for life test is: select n products as a sample to test, when test cumulative time reaches to the expected value T, stop the test, Assume that r times of failure happened in the process of test, if $r \leq c$, we think that products are qualified, then accept the original hypothesis, if $r > c$, then, accept the alternative hypothesis, products are rejected. So, the main task of time censoring test design is to choose appropriate c and T.

4.2 The Plan Based on Average Risk Criteria

As well as failure censoring plan, according to the Bayes theory, take the conjugate prior distribution for the Gamma distribution, which is written as G (a, b), that is:

$$\pi(\lambda|a,b) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda} \tag{18}$$

Among, $\Gamma(a)$ is the Gamma function, defined as:

$$\Gamma(a) = \int_0^\infty e^{-x}x^{a-1}dx \tag{19}$$

According to cumulative distribution function $F(t) = 1-\exp(-t/\theta)$ of exponential distribution, the reliability of the product is $R(t) = \mathrm{e}^{-\lambda t}$, the probability that r failures occurred in n products until time t is:

$$\binom{n}{r}F(t)^r R(t)^{n-r} \tag{20}$$

The failure rate of products $r \leq c$ until time t, and probability that product is accepted is

$$L(\lambda) = \sum_{r=0}^{c}\binom{n}{r}F(t)^r R(t)^{n-r} \tag{21}$$

Because the value of λ is small, so we expand $R(t) = \mathrm{e}^{-\lambda t}$ by using Taylor formula:

$$\begin{aligned} R(t) &= e^{-\lambda t} = 1 - \lambda t + \frac{1}{2!}\lambda^2 t^2 - \cdots \approx 1 - \lambda t \\ F(t) &= 1 - R(t) = \lambda t \end{aligned} \tag{22}$$

Then, the probability of acceptance is:

$$L(\lambda) = \sum_{r=0}^{c} \binom{n}{r} (\lambda t)^r (1 - \lambda t)^{n-r} \tag{23}$$

Under the condition of $n\lambda t \leq 5$, $F(t) \leq 10\ \%$, binomial probability can be available by poison probability approximately, so we can obtain that:

$$L(\lambda) = \sum_{r=0}^{c} e^{-n\lambda t} \frac{(n\lambda t)^r}{r!} \tag{24}$$

Generally, n is smaller, so $T \approx nt$, then

$$L(\lambda) = \sum_{r=0}^{c} e^{-\lambda T} \frac{(\lambda T)^r}{r!} \tag{25}$$

Based on average risk criteria and the Bayes formula, Combined with the prior distribution of the failure rate λ and the expression for probability of acceptance, we can deduce the calculation formula of two types of risk, they are as follows:

$$\alpha(c, T) = P(t < T | \lambda < \lambda_0) = \frac{\int_0^{\lambda_0} (1 - \sum_{r=0}^{c} e^{-\lambda T} \frac{(\lambda T)^r}{r!}) \pi(\lambda) d\lambda}{\int_0^{\lambda_0} \pi(\lambda) d\lambda} \tag{26}$$

$$\beta(c, T) = P(t \geq T | \lambda > \lambda_1) = \frac{\int_{\lambda_1}^{\infty} \sum_{r=0}^{c} e^{-\lambda T} \frac{(\lambda T)^r}{r!} \cdot \pi(\lambda) d\lambda}{\int_{\lambda_1}^{\infty} \pi(\lambda) d\lambda} \tag{27}$$

Test plan can be getting by solving the two equations.

4.3 The Plan Based on Posteriori Risk Criteria

According to the posteriori risk criteria, and the derivation for prior distribution of exponential products and the expression for probability of acceptance in the previous section, the calculation formula for the first category of risk $\alpha(r, T)$ is:

$$\alpha(c,T) = P(\lambda \le \lambda_0 | t < T) = \int_0^{\lambda_0} p(\lambda | t < T) d\lambda$$
$$= \frac{\int_0^{\lambda_0} P(t < T|\lambda)\pi(\lambda)d\lambda}{\int_0^{\infty} P(t < T|\lambda)\pi(\lambda)d\lambda} = \frac{\int_0^{\lambda_0} (1 - \sum_{r=0}^{c} e^{-\lambda T} \frac{(\lambda T)^r}{r!})\pi(\lambda)d\lambda}{\int_0^{\infty} (1 - \sum_{r=0}^{c} e^{-\lambda T} \frac{(\lambda T)^r}{r!})\pi(\lambda)d\lambda} \quad (28)$$

The calculation formula for the second category of risk $\beta(r, T)$ is:

$$\beta(r,T) = P(\lambda > \lambda_1 | t \ge T) = \int_{\lambda_1}^{\infty} p(\lambda | t \ge T) d\lambda$$
$$= \frac{\int_{\lambda_1}^{\infty} P(t \ge T|\lambda)\pi(\lambda)d\lambda}{\int_0^{\infty} P(t \ge T|\lambda)\pi(\lambda)d\lambda} = \frac{\int_{\lambda_1}^{\infty} (1 - \sum_{r=0}^{c} e^{-\lambda T} \frac{(\lambda T)^r}{r!})\pi(\lambda)d\lambda}{\int_0^{\infty} (1 - \sum_{r=0}^{c} e^{-\lambda T} \frac{(\lambda T)^r}{r!})\pi(\lambda)d\lambda} \quad (29)$$

5 Comparison of Plans Under Two Criteria and the Effect of Prior Distribution

In this section, through a simple example to illustrate, assume that we design the life verification testing of product under exponential distribution, and we know the priori distribution of λ that is G(4, 725),and we require that the upper limit of inspection is not more than 0.003, then design the life verification testing of product based on average risk criteria and posteriori risk criteria, take life verification test under failure censoring for example, plan can be obtained respectively and comparison with classical solutions [7] was shown in the following table by combining (12), (13), (14) and (15) (Table 1).

Table 1 Two plans and comparison with classical solutions

Failure censoring	$\lambda_1 = 0.003$, $\theta_0 = 333$ h, $\theta_1 = 166$ h, discrimination ratio $d = \theta_0/\theta_1 = 2$							
Both risk	$\alpha = 0.05$, $\beta = 0.05$		$\alpha = 0.05$, $\beta = 0.1$		$\alpha = 0.1$, $\beta = 0.05$		$\alpha = 0.1$, $\beta = 0.1$	
Plan	r	T	r	T	r	T	r	T
Classical solutions	23	5,215	19	4,130	18	4,250	15	3,420
Average risk criteria	9	2,000	7	1,400	7	1,400	5	1,100
Posterior risk criteria	5	1,400	4	1,000	4	1,000	2	750

Note The priori distribution of Bayes plan is gamma distribution, G (4, 725).

It can be seen that using Bayes method could largely save the test time and reduce the cost by comparing the three plans, but there are still some differences in the plan of average risk criteria and posterior risk criteria, take the plan that both risks is 0.05 as an example, through changing the parameter b in the priori distribution of λ, we can get the change of critical time under the two criteria, as shown below (Figs. 1 and 2):

From the two figures we know that prior distribution has less effect on the plan under the average risk criteria, but more effect on the posterior risk criteria's one, this also proves the difference between the two criteria mentioned in Sect. 2.3.

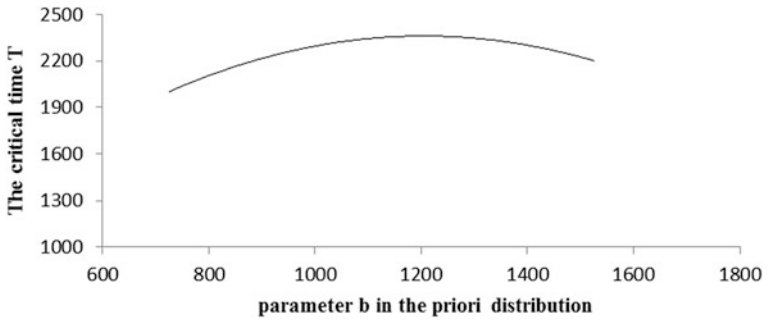


Fig. 1 The effect of priori distribution on plan under the average risk criteria

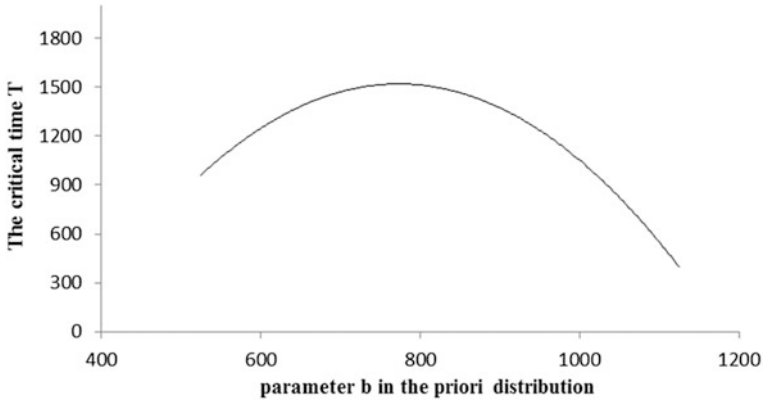


Fig. 2 The effect of priori distribution on plan under the posterior risk criteria

6 Conclusion

This paper aims at the two kinds of censoring method (type-I and type-II censoring), we give the Bayesian acceptance plan of the producer and consumer risks according to the Average risk criteria and Posterior risk criteria respectively, and the test plan of different risk criteria as well as the classical methods are compared, the results show that the application of Bayes plan can greatly reduce the test cost. the influence of the prior distribution on the Bayes plan is studied, it show that prior distribution has more effect on the plan under the average risk criteria, this remind us that choosing the plan according to the actual situation.

References

1. Balakrishnan N, Leiva V, López J (2007) Acceptance sampling Plans from truncated life tests based on the generalized Birnbaum-Saunders Distribution. Commun Stat Simul Comput 36 (3):643–656
2. Liang TC, Yang M-C (2011): Optimal Bayesian sampling plans for exponential distributions based on hybrid censored samples. J Stat Comput Simul doi:10.1080/00949655.2011.642378
3. Amin Z, Salem M (2012) On designing an acceptance sampling plan for the Pareto lifetime model. J Stat Comput Simul 82(8):1115–1133
4. Lin C-T, Huang Y-L, Balakrishnan N (2008) Exact Bayesian variable sampling plans for the exponential distribution based on type-I and type-II hybrid censored samples. Commun Stat Simul Comput 37(6):1101–1116
5. Fallah Nezhad MS, Hosseini Nasab H (2012) A new Bayesian acceptance sampling plan considering inspection errors. Sharif University of Technology, Tehran
6. Hamada MS, Wilson Alyson G, Shane RC (2008) Bayesian reliability. Springer, New York
7. Jiang T (2012) Reliability and life test. National Defense Industry Press, Beijing, pp 225–226

An Approach Based on Frequency Domain for Random Vibration Fatigue Life Estimation

Jing Hailong, Chen Yunxia and Kang Rui

Abstract According to the structural vibration fatigue characters, a life assessment method is presented. The method is based on the information of the frequency domain of random. Firstly, the frequency response of the structure analysis should be carried out by finite element analysis (FEA) under random vibration load, and the position and the stress response power spectral density (PSD) function can be obtained. Secondly, the life calculation model for the single-stage load is provided based on RC and LCC, and the cumulative damage for the structure on the multi-stage load can be obtained using Miner linear theory. Then the fatigue life of structure can be estimated under random vibration load. Lastly, a case of the life estimation method is presented.

Keywords Vibration fatigue · Life estimation · FEA · PSD

1 Introduction

The baseband modal response is only considered for the life assessment using the stress response spectrum in the traditional engineering application methods. Scilicet, the problem reduces to the ideal narrow-band process whose center frequency is the first-order natural frequency of the structure. However, this simplified engineering method ignores the possible existence of high order modes effect of normal stress response components on structural fatigue damage, which may result in insufficient damage estimates and high life estimation [1]. If effects of high order

J. Hailong (✉) · C. Yunxia · K. Rui
School of Reliability and Systems Engineering, BUAA, Beijing, China
e-mail: Jinghailong@dse.buaa.edu.cn

C. Yunxia
e-mail: chenyunxia@buaa.edu.cn

K. Rui
e-mail: kangrui@buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_121

modes normal stress response are considered for the fatigue damage estimation of structures, the problem becomes a study of a kind of special consists of multiple narrow-band broadband issues.

For broadband process, there are many common methods for life estimation such as the correction factor method based on narrow-band hypothesis proposed by Wirsching [2, 3], Dirlik [4, 5] and so on. Besides, Bi-modal approach [6] is proposed by Fu for the special broadband process constituted by the bimodal response, and this approach is good for the stress response problems whose modal frequency ratio greater than or equal to 4.0 dual modal (or dual narrowband). However, the calculating formulas are complex and the calculating process is tedious for this method.

For broadband process consisted by multi modal responses, this article presents a method for fatigue life estimation based on rain-flow counting method for random vibration. The method is based on the information of the frequency domain of random. Firstly, vibration simulation analysis of structures under random vibration loads are conducted using Workbench finite element analysis software and the dangerous point of the stress response and spectral density of the structure for the per stage load are obtained under the random excitation load. Secondly, the damage of the per stage load can be calculated used the life calculating model based on RC and LCC, and the cumulative damage for the structure on the multi-stage load can be obtained using Miner linear theory. Lastly, the life can be estimated and a case of the life evaluation method is presented.

2 The Life Estimation Steps

Firstly, the dynamic response of structures under random vibration load can be obtained through the random vibration theory and the finite element analysis. Secondly, the dynamic response of structures under random vibration load of single-stage can be got using the life calculating model based on RC and LCC. Then, the cumulative damage for the structure on the multi-stage load can be obtained using Miner linear theory and the life can be estimated. The life calculated specific steps are shown in Fig. 1.

3 The Random Vibration Analysis

3.1 The Method of the Random Vibration Analysis

The random vibration analysis is the method which is used to analyze the response characteristics of structures or components under random load using probabilistic and statistical methods. With the known power spectral density of random load, the statistical properties of the random responses of structures or parts can be obtained through the establishment of the relationship for structures or components between the input and output spectrum function.

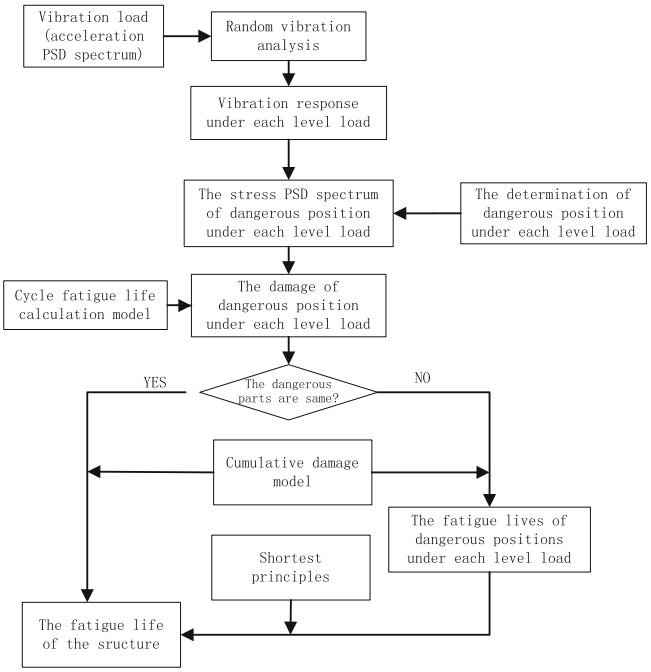


Fig. 1 The life calculated specific steps

The random vibration response can be obtained through following two a way: first, the units pulse response function will be got from Shi domain strike system and the response for the coefficient linear system on any incentive can be calculated through the time integral for the convolution of this function and the excitation functions; second, the frequency response function will be got from frequency domain strike system and the system frequency response function can be calculated through the matrix operations between the incentive power spectrum and the system frequency response function. The second method can be used for the problem that the information for Shi domain strike system cannot be obtained, and the load power spectrum time domain information is known, while the first method cannot be used. The relationship between the load power spectral density $S_F(\omega)$ and the response power spectrum density $S_X(\omega)$ is as follows for single excitations:

$$S_X(\omega) = |H(\omega)|^2 S_F(\omega) \tag{1}$$

The structural response variance is:

$$\sigma_X^2 = \int_{-\infty}^{+\infty} S_X(\omega) d\omega = \int_{-\infty}^{+\infty} |H(\omega)|^2 S_F(\omega) d\omega \tag{2}$$

The relationship of power spectral density between load and response are as follows:

Speed power spectral density

$$S_{\dot{X}}(\omega) = \omega^2 |H(\omega)|^2 S_F(\omega) \tag{3}$$

Acceleration response power spectrum density

$$S_{\ddot{x}}(\omega) = \omega^4 |H(\omega)|^2 S_F(\omega) \tag{4}$$

If more than one incentive, such as $F_1(\omega), F_2(\omega), F_3(\omega), \cdots$, are independent each other, the incentive system response power spectrum $S(\omega)$is the total response power spectrum density.

$$S(\omega) = \sum S_{Xi}(\omega) = \sum |H(\omega)|^2 S_{Fi}(\omega) \tag{5}$$

Where, power spectral density for the response is the Fourier transform of the response-related function.

The power spectral density for the response is

$$S_X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} R_X(\tau) e^{-i\omega\tau} d\tau \tag{6}$$

The response-related function is

$$R_X(\tau) = E[X(t)X(t+\tau)] \tag{7}$$

The transfer function is

$$H(\omega) = \frac{1}{\omega_0^2 - \omega^2 + i2\xi\omega_0\omega} \tag{8}$$

Therefore, the response statistical characteristics of structures or components can be determined by random vibration analysis with the random load spectrum known.

3.2 The Process and Methodology

The random vibration analysis is the random vibration spectrum analysis based on probability statistics. The key is to obtain the transfer function matrix and the statistics characteristics of structure response can be calculated with the power spectral density of random vibration excitation known. The transfer function usually is obtained using modal superposition method and the workbench finite

element analysis software is used to get the random vibration response. Thus, the vibration stress analysis processes are as Fig. 2, and the processes can be divided four steps, such as the establishment of digital prototyping solution calculation models, the calculation of modal solutions, the calculation of spectrum solutions, the result fetch.

4 The Fatigue Life Calculation

4.1 The Life Calculation Model of the Cycle Counting Method

The cycle counting method is divided three steps: first, the amplitude information is available through cycle counting the stress peak probability density function; second, the structure damage expectations are calculated using the theory of cumulative damage; last, the fatigue life of the structure can be estimated.

Frequency-domain method to calculate the cumulative damage is generally based on the following formula [7]:

$$E(D) = vaC \int_0^{+\infty} s^k pa(s) ds \tag{9}$$

where, $E(D)$ is the expectation of the damage for the per unit of time; va is the cycle probability for load (the number of cycles of load appearing in per unit of time); C is the constant in the S-N curve ($s^k N = \frac{1}{C}$); $pa(s)$is the probability distribution of the stress amplitude.

Peak counting method (PC), level crossing notation (LCC), change-counting method (RC) and rain-flow cycle notation (RFC) are four cycle notation that are the most commonly used. The relationships among the expectations for fatigue damage under four counting are given by Frendahl and Rychlik [8].

$$E\left(D^{RC}\right) \le E\left(D^{RFC}\right) \le E\left(D^{LCC}\right) \le E\left(D^{PC}\right)$$

The rain-flow counting loops are the most commonly used method of counting, but the process of rain-flow cycle count is cumbersome and the accurate expression is difficult to resolve [9, 10].The rain-flow counting method can be approximate a linear combination of the law between count-counting method and level crossing, which is proposed by Tovo [11] based on the above inequality.

$$E\left(D^{RFC}\right) = b\ E\left(D^{LCC}\right) + (1-b)\, E\left(D^{RC}\right) \tag{10}$$

where, b is between 0 and 1 and there is not a theoretical method to the value of b. The formula is proposed through numerical simulation based on probability

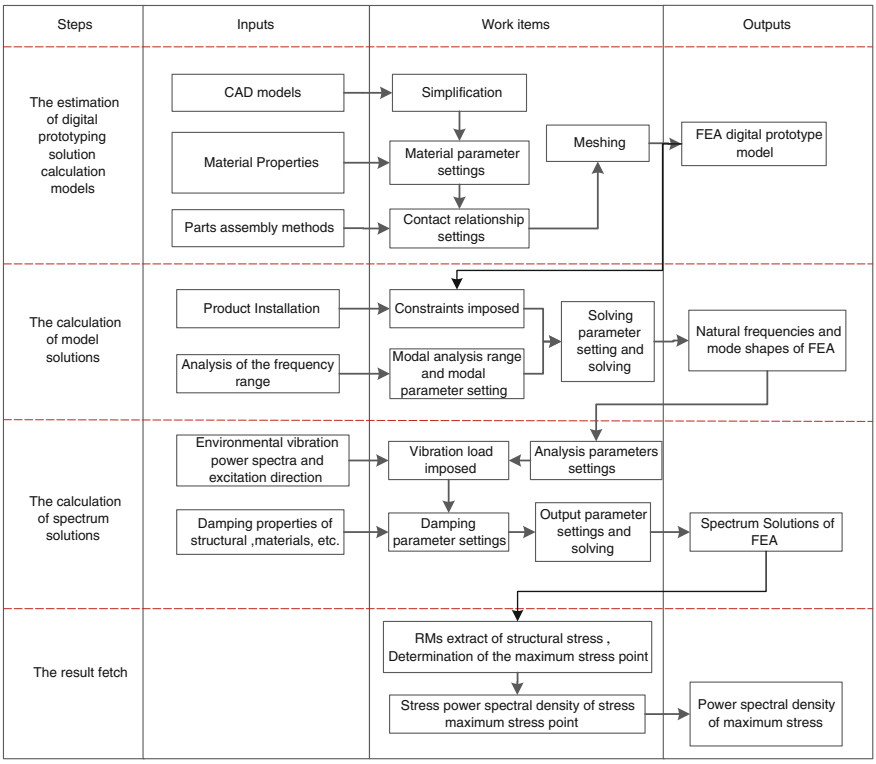


Fig. 2 The random vibration analysis processes

density approximation of spectral moments of peak expression by Roberto Tovo and Benasciutti. The formula [12, 13] of the value is as follow:

$$b = \frac{(\alpha_1 - \alpha_2)[1.112(1 + \alpha_1\alpha_2 - (\alpha_1 + \alpha_2))e^{2.11\alpha_2} + (\alpha_1 - \alpha_2)]}{(\alpha_2 - 1)^2} \tag{11}$$

The approximate formula of fatigue damage [4] is proposed by Benasciutti and Tovo based on change-counting method (RC) and rain-flow cycle notation (RFC). The formulas are as follows:

$$E(D^{LCC}) = vpC\alpha_2(\sqrt{2}\sigma x)^k \Gamma(1 + \frac{k}{2}) \tag{12}$$

$$E(D^{RC}) \cong E(D^{LCC})\alpha_2^{k-1} \tag{13}$$

where,

vp is the peak probability: $vp = \frac{1}{2\pi}\sqrt{\frac{\lambda 4}{\lambda 2}}$

α_1 and α_2 are the bandwidth parameters: $\alpha_2 = \frac{\lambda_2}{\sqrt{\lambda_0 \lambda_4}}$, $\alpha_1 = \frac{\lambda_1}{\sqrt{\lambda_0 \lambda_2}}$

σ_x is the Standard deviation $\sigma_x = \sqrt{\lambda_0}$

4.2 The Calculation of the Damage Cumulative Life

At present, there is no specific vibration fatigue damage cumulative theory. Some experts pointed out that, in view of the dynamic fatigue estimation error is large, other relevant non-linear theory of cumulative damage does not significantly improve analysis accuracy, but increased the workload and difficulty of analysis [14].Thus, the Miner linear accumulate damage are used to calculate the cumulative damage in this paper.

$$D = \sum_{i=1}^{p} D_i \tag{14}$$

Miner Assumes that when the amount of damage (D) equals 1, the specimen fatigue failure occurs. However, this norm is relatively conservative during the process, in application, especially in random vibration [15]. In the practical engineering, the specimen fatigue failure of the structure occurs, which bears the high–low load, when D < 1; while the specimen fatigue failure of the structure occurs, which bears the low–high load, when D > 1. Obviously, there is a close relationship between the amount of damage and loaded order. In addition, hang Yao pointed out that the value of D for sinusoidal vibration is desirable to 1–1.5 and the value of D for random vibration is desirable to 1.5–2. If sinusoidal and random vibrations exist simultaneously, it access to 1.5 [16].

The life of structures under random vibration load in comprehensive damage is as follows:

$$L = \frac{1}{\sum_{i=1}^{p} D_i} \tag{15}$$

5 The Examples for Application

The imported fitting in an engine-driven pump is taken on as a example to show the process of life estimation for the structure under the random vibration.

5.1 Random Vibration Analysis

(1) The establishment of digital prototyping solution calculation models

The FEA model is estimated by workbench for the imported fitting. Figure shows the model of the imported fitting (Fig. 3).

(2) Modal analysis

The dynamic characteristics of the main model should be known in the model analysis to get each order natural frequency in the frequency bands covered by the PSD Load spectrum. The result for the top 15 order modal is showed in Fig. 4.

(3) Frequency response analysis

It is necessary for frequency response analysis that the acceleration power spectra in the entire band should be imposed on the inspiring direction as the excitation spectra (showed in Fig. 5). The transfer function of the structure is obtained through the frequency response analysis by applying load excitation spectrum on the structure and the stress response distribution for the 15—order

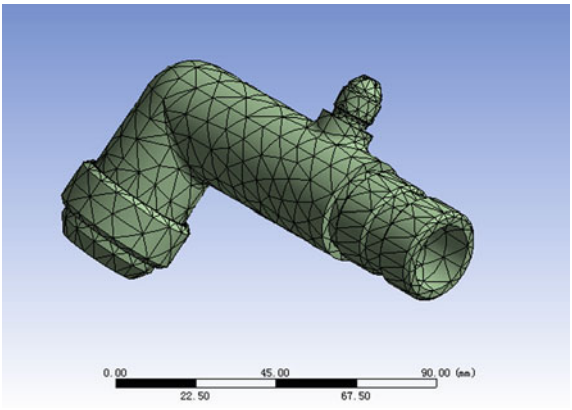


Fig. 3 The model of the imported fitting

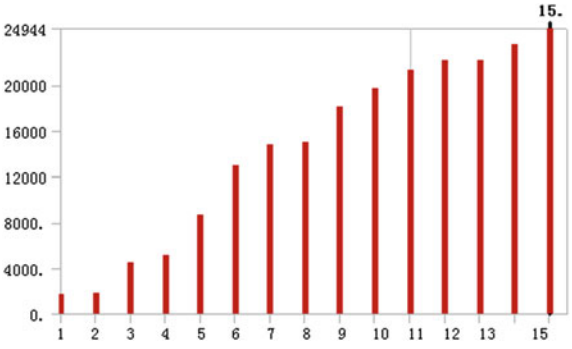


Fig. 4 The top 15 order model of the structure

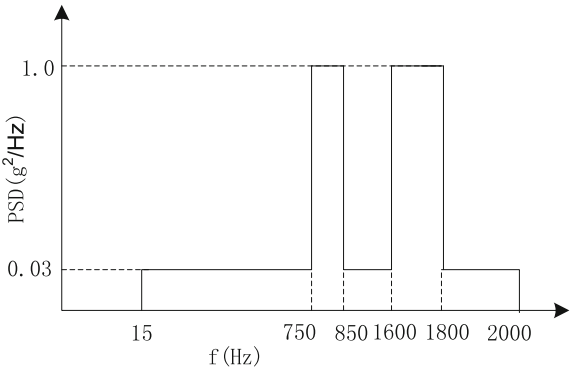


Fig. 5 The acceleration power spectra

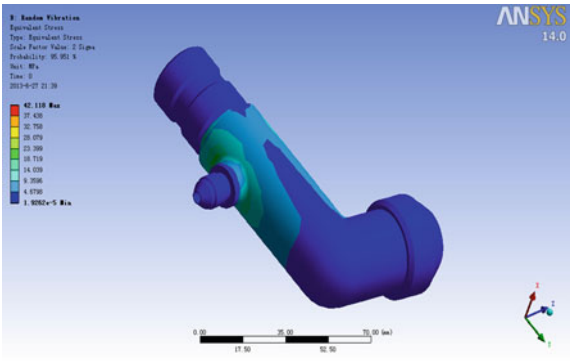


Fig. 6 The stress response distribution

model under single-stage load is showed in Fig. 6. The position of the point of the maximum stress (weak position) is showed in the figure of the stress response distribution and the stress PSD distribution can be gotten by extracting the stress response distribution of the weakest location. The stress PSD distribution is showed in Fig. 7.

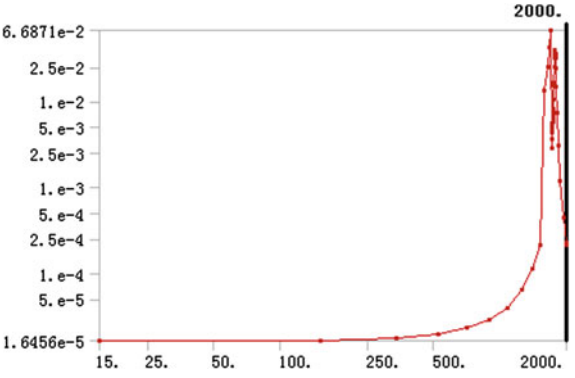


Fig. 7 The stress PSD distribution of the weakest location

5.2 The Fatigue Life Calculation

The fatigue life under the single-stage load can be calculated as 17,849 h by the life calculation model of the cycle counting method according to the abscissa and ordinate values of the stress PSD of the point of maximum stress.

The acceleration PSD at different stages may be different throughout the whole life cycle. Thus, the damages of all weak points at different stages need to be calculated, and the cumulative damage will be accumulated though the Miner linear accumulate damage model. Lastly, the fatigue vibration life under the frequency domain can be obtained according to the shortest possible principles.

6 Conclusion

With the engineering Simulation Software of Workbench, this paper presents a method for fatigue life estimation based on rain-flow counting method for random vibration and can give a quantitative result for predicting the fatigue life under random vibration. The method resolves the blank for quantitative estimation for the fatigue life under random vibration and provides the bridge between the academic and the industry.

References

1. Cao M, Shao C, Qi P (2011) The test verification for the fatigue life frequency-domain estimation method under broadband random vibration. Res Struct Strength 11(1):29–34
2. NWM Bishop (1999) Vibration fatigue analysis in the finite element environment. In: XVI Encuentro del grupo espanol de fractura Torremolinos,Spain, 14–16 April 1999
3. Chow CL, Li DL (1991) An analytical solution for fast fatigue assessment under wide-band random load[J]. Int J Fatigue 13(5):395–404
4. Benasciutti D, Tovo R (2006) Comparison of spectral methods for fatigue analysis of broad-band Gaussian random processes. Probab Eng Mech 21(4):287–299
5. Wirsching PH, Torng TY (1991) Advanced fatigue reliability analysis. Int J Fatigue 13 (No5):389–394
6. Fu TT, Cebon D (2000) Predicting fatigue lives for Bi-modal stress spectral densities. In J Fatigue 22:11–21
7. Rychlik I (1993) Note on cycle counts in irregular loads. Fatigue Fract Eng Mater Struct 16 (4):377–390
8. Frendahl M, Rychlik I (1993) Rainflow analysis: Markov method. Int J Fatigue 15(4):265–272
9. Petrucci G, Zuccarello B (1999) On the estimation of the fatigue cycle distribution from spectral density data. J Mech Eng Sci 213(8):819–831
10. Petrucci G, Di Paola M, Zuccarello B (2000) On the characterization of dynamic properties of random Processes by spectral parameters. J Appl Mech 67(3):519–526
11. Tovo R (2002) Cycle distribution and fatigue damage under broad- band random load. Int J Fatigue 24(11):1137–1147

12. Benasciutti D, Tovo R (2005) Spectral methods for lifetime prediction under wide-band stationary random processes. Int J Fatigue 27(8):867–877
13. Benasciutti D, Tovo R (2005) Cycle distribution and fatigue damage assessment in broad-band non-Gaussian random processes. Probab Eng Mech 20(2):115–127
14. Qihang Yao, Jun Yao (2006) Vibration fatigue problems of engineering structures. J Appl Mech 23(1):12–15
15. Xiaohua Yang, Weixing Yao, Chengmei Duan (2003) Deterministic fatigue cumulative Damage. China Eng Sci 5(4):81–87
16. Yao Q, Yao J (2009) Structural vibration fatigue characteristics and analytical methods. Mech Sci Technol 19:9

Research on the Wear Life Analysis of Aerohydraulic Spool Valve Based on a Dynamic Wear Model

Liao Xun, Chen Yunxia and Kang Rui

Abstract Leakage due to wear is one of the main failure modes of aero-hydraulic spool valve. This paper established a dynamic wear model based on dynamic system modelling theory. Firstly, the wear mechanism between spool and valve sleeve of aero-hydraulic valve is analysed followed by the verification through observation of the wear morphology of the spool surface by SEM. Secondly, the dynamic wear model is established from three aspects of fractional film defect index, Rms of profile, adhesive wear coefficient based on Archard wear model. Lastly, some qualitative analysis on the wear regularity of aero-hydraulic spool valve is conducted based on the dynamic wear model established.

Keywords Adhesive wear · Fractional film defect index · Adhesive wear coefficient · Rms of profile · Dynamic wear model

1 Introduction

Aero-hydraulic spool valve is the controlling component of the aircraft hydraulic system whose working condition will impose a direct effect on the performance of the hydraulic system. Leakage due to wear is one of the main failure modes of aero-hydraulic spool valve. Most of the current research on aero-hydraulic spool valve focuses on the loss of function due to hydraulic clamping and pollution clamping [1, 2]. But related research on the performance degradation due to wear between spool and valve sleeve is not much [3].

L. Xun (✉) · C. Yunxia · K. Rui
Beijing University of Aeronautics and Astronautics, Beijing, China
e-mail: liaoxun@126.com

C. Yunxia
e-mail: chenyunxia@buaa.edu.cn

K. Rui
e-mail: kangrui@buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_122

This paper established a dynamic wear model of aero-hydraulic spool valve based on dynamic system modelling theory and modified Archard wear model. Fractional film defect index β is introduced into the Archard wear model to modify the real metal-to-metal contact area under mixed lubrication condition of spool and valve sleeve and the quantitative formula of β is derived. The quantitative relation of the variation of the Rms of profile σ_s which is a characterization of surface roughness with time is also derived based on the truncating process of asperities during wear. The quantitative correlation between the wear coefficient K_{adh} of Archard wear model and σ_s is also derived in order to describe the dynamic characteristic of the change of surface roughness on wear.

2 Analysis and Verification of Wear Mechanism

2.1 Wear Mechanism Analysis

The wear mechanism analysing process takes into consideration all the input factors of the tribological system composed of spool and valve sleeve which includes material properties, working condition and load condition as well as all the possible variations of system properties.

The spool and valve sleeve of aero-hydraulic valve is made of the same material. Due to the macro and micro geometrical defects on the outer surface of spool and inner surface of valve sleeve caused by the machining and assembly process, many micro peaks or asperities and valleys still remain on the outer surface of spool and inner surface of valve sleeve which is much bigger compared with the size of metal atoms. Due to the small designed coordinate interval between spool and valve sleeve, the asperities on both surfaces will contact and further extrude and deform during the relative motion between spool and valve sleeve. The high contact pressure on the contact point will cause the adhesion of metals which will further lead to the adhesive wear between spool and valve sleeve.

2.2 Wear Mechanism Verification

The wear mechanism verification is based on the observation of the wear morphology through SEM on the same shoulder of the spool by the comparison of the initial state and the state after 10000 conversions. The aero-hydraulic spool valve observed is presented in Fig. 1.

The SEM results of the initial state of the shoulder observed is shown in Fig. 2 and the results after 10000 conversions is shown in Fig. 3. Shedding and transformation of material due to the shear fracture of adhesion points can be seen in Fig. 3 which can work as a proof of the adhesion wear mechanism between spool and valve sleeve.

Fig. 1 Aero-hydraulic spool valve

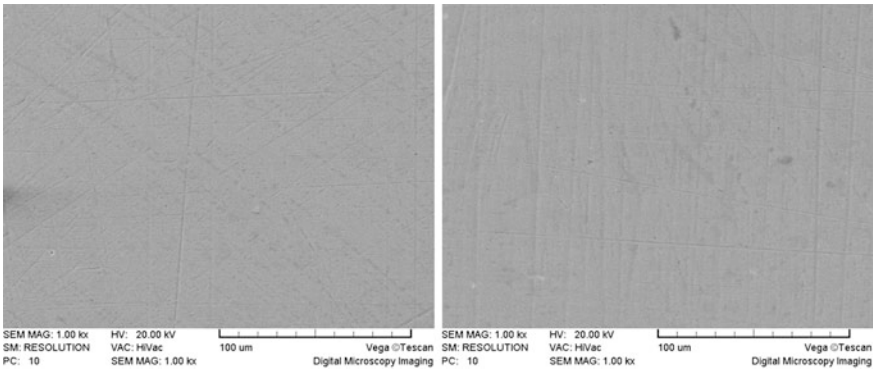
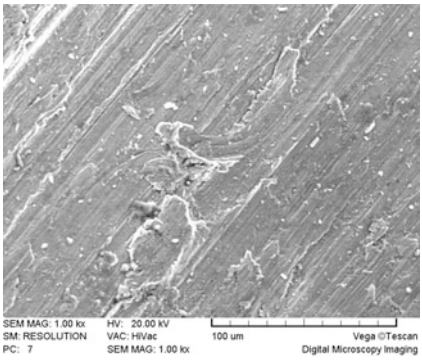


Fig. 2 SEM results of initial state

Fig. 3 SEM results after 10,000 conversions



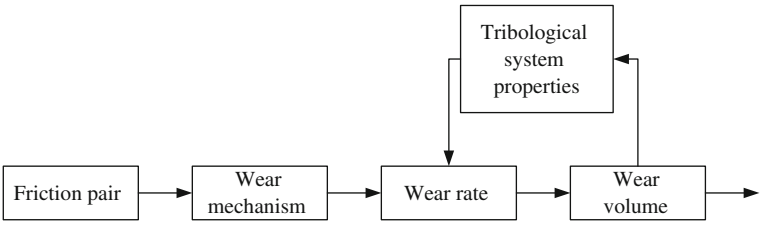


Fig. 4 Single variable first order tribological system

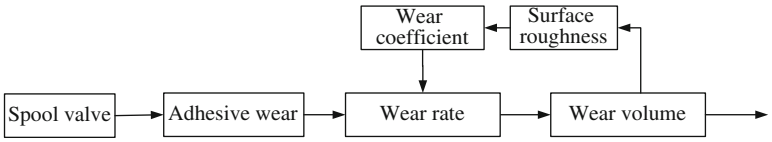


Fig. 5 Double variable first order tribological system of spool valve

3 Dynamic Wear Modelling

Tribological system is a system composed of tribological elements which are abstracted from elements in mechanical system or other natural system and is often used to analyse the behaviour and results of tribological elements. The geometrical, physical and chemical properties of a tribological system is in continuous change during wear process, so dynamic system modelling theory can be applied in quantitative wear modelling. Differential equations are often used in dynamic system modelling to describe the non-stationary and steady state of the system.

If only one independent state variable of the tribological system is considered, the feedback loop is shown in Fig. 4 [4].

The author takes surface roughness and wear volume as two independent state variable of the corresponding tribological system of spool valve which leads to a double variable first order tribological system whose feedback loop is shown in Fig. 5.

3.1 Modified Archard Wear Model

Archard wear model is widely applied in the calculation of adhesive wear under dry condition. It was first proposed by Archard [5] which takes the form

$$V = K\frac{F_N}{H}L \tag{1}$$

where V is the wear volume, K is the wear coefficient, F_N is the normal load, H is the material hardness and L is the wear stroke.

In the case of aero-hydraulic spool valve, the normal load between spool and valve sleeve is the asperity load due to the coordinate interval, so we change the normal load F_N into asperity load W_a. Also, aero-hydraulic spool valve works in mixed lubrication condition where both oil film support and metal-to-metal contact exist. While adhesive wear only takes place in metal-to-metal contact points, so it is reasonable to modify the Archard wear model when the effect of oil film is taken into consideration. The fractional film defect index β is introduced into the Archard wear model to modify the real metal-to-metal contact area which leads to Eq. (2).

$$V = K\beta\frac{W_a}{H}L \tag{2}$$

3.2 Fractional Film Defect Index β

The fractional film defect index β is used to characterize the reduction of wear by lubrication oil whose definition is as follows

$$\beta = \frac{A_m}{A_r} \tag{3}$$

where A_r is the contour contact area which includes oil film contact area and metal-to-metal contact area, A_m is the metal-to-metal contact area. Fractional film defect index β is determined by the adsorption and desorption capacity of oil molecules on the surface of friction pair. Factors that affect the adsorption and desorption capacity of oil molecules on the surface of friction pair during the relative motion process of friction pair is the oil temperature and relative motion velocity.

Kingsbury [6] has presented the calculation formula of fractional film defect index β under single lubricant condition

$$1 - \beta = \exp\left(-\frac{t_z}{t_r}\right) \tag{4}$$

where t_z is the time consumed by the asperity slides through the equivalent distance of oil molecular diameter, t_r is the average adsorption time of an oil molecular on a certain point of the friction pair surface. The time t_z can be calculated by Eq. (5)

$$t_z = \frac{D}{v} \tag{5}$$

where D is the oil molecular diameter in adsorption state, v is the relative motion velocity of friction pair. If the oil molecular is taken as a sphere, then D can be calculated by Eq. (6)

$$D = \left(\frac{6V_m}{\pi N_a}\right)^{\frac{1}{3}} \tag{6}$$

where V_m is the molar volume of oil molecular, N_a is the Avogadro constant and if we take N_a as 6.02×10^{23}, Eq. (6) can be transformed into Eq. (7).

$$D = 1.4 \times 10^{-8} V_m^{\frac{1}{3}} \tag{7}$$

Frenkel [6] provided the calculation formula of t_r

$$t_r = t_0 \exp\left(\frac{E_c}{RT_s}\right) \tag{8}$$

where t_0 is the oscillation period of an adsorbed oil molecular on the surface of friction pair, E_c is the adsorption energy of oil molecular, T is the oil temperature, R is the Gas constant.

Combining Eqs. (4), (5), (7) and (8) leads to the calculation formula of fractional film defect index β.

$$\beta = 1 - \exp\left\{-\left[\frac{1.4 \times 10^{-8} V_m^{\frac{1}{3}}}{v t_0}\right] \exp\left(-\frac{E_c}{RT}\right)\right\} \tag{9}$$

3.3 Rms of Profile σ_s

Surface roughness of both spool and valve sleeve will change with time due to the truncating process caused by the extrusion and collision of asperities on both surfaces during the wear process. Rms of profile σ_s is often chosen to characterise the geometrical properties of mating surfaces during wear including surface roughness. The contact of two randomly rough surfaces of spool and valve sleeve are equivalently transformed into the contact of a rigid smooth surface and a composite elastoplastic rough surface [7] as shown in Fig. 6.

As the wear proceeds, asperities of the composite elastoplastic rough surface will be truncated which will lead to the variation of the height distribution function of surface profile with time as shown in Fig. 7.

Suppose the possibility density function of the profile Z of composite rough surface at time t is φ(z,t) where φ(z,0) is known as the initial state possibility density function. We define h the distance between the profile centerline of two mating surfaces (Fig. 6) and w(z-h) as the asperity wear velocity function which stands for the reduction of height per unit time of asperities at height z or asperities having a embed depth of (z-h). According to reference [8], we have

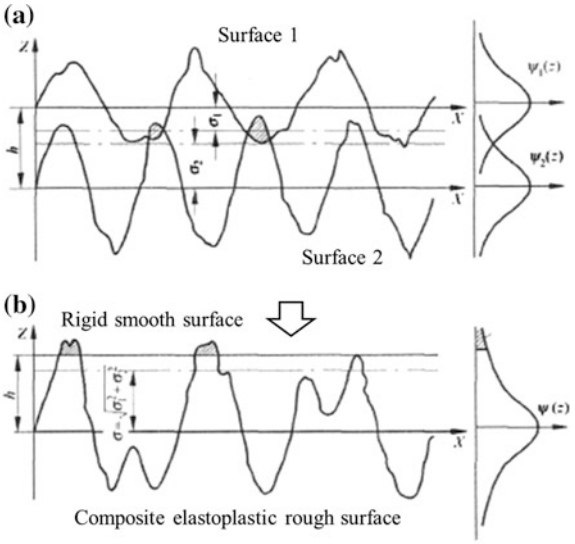


Fig. 6 Equivalent transform of two randomly roughness mating surfaces

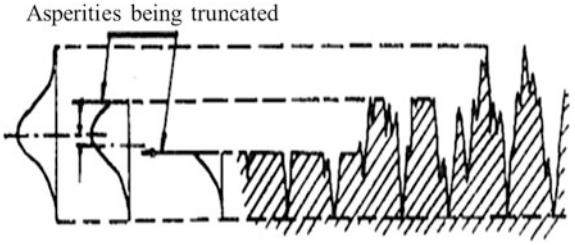


Fig. 7 Asperity truncating process during wear

$$\frac{\partial}{\partial t}\varphi(z,t) = \frac{\partial}{\partial z}[w(z-h)\varphi(z,t)] \tag{10}$$

The Rms of profile of composite rough surface σ_s is

$$\sigma_s^2 = E[Z^2(t)] - \{E[Z(t)]\}^2 = \int_{-\infty}^{+\infty} z^2\varphi(z,t)dz - \left[\int_{-\infty}^{+\infty} z\varphi(z,t)dz\right]^2 \tag{11}$$

Take the derivative with respect to t on both sides of Eq. (11)

$$2\sigma_s\frac{d\sigma_s}{dt} = \int_{-\infty}^{+\infty}\left[2z\cdot\frac{dz}{dt}\cdot\varphi(z,t) + z^2\cdot\frac{\partial}{\partial t}\varphi(z,t)\right]dz + 2\int_{-\infty}^{+\infty} z\varphi(z,t)dz \cdot \{\int_{-\infty}^{+\infty}\left[\frac{dz}{dt}\cdot\varphi(z,t) + z\cdot\frac{\partial}{\partial t}\varphi(z,t)\right]dz\} \tag{12}$$

Substitute Eq. (10) into Eq. (12)

$$\frac{d\sigma_s}{dt} = -\frac{2}{\sigma_s}\left[\int_{-\infty}^{+\infty} z \cdot w(z-h) \cdot \varphi(z,t)dz + \int_{-\infty}^{+\infty} z \cdot \varphi(z,t)dz \cdot \int_{-\infty}^{+\infty} w(z-h) \cdot \varphi(z,t)dz\right] \tag{13}$$

Suppose the asperity wear velocity function w(z-h) takes the following form

$$w(z-h) = \begin{cases} D(z-h)^{\gamma}, \\ 0, \end{cases} \begin{matrix} , & z > h \\ , & z < h \end{matrix} \tag{14}$$

where γ is an exponent characterizing the degree of asperity deformation, D is a coefficient associated with the wear rate of the friction pair, and we have [4]

$$D = C \cdot \frac{dx}{dt} \tag{15}$$

where the coefficient C > 0, x is the wear depth. Consider the relationship between volume wear rate and depth wear rate

$$\frac{dV}{dt} = A_a \frac{dx}{dt} = K\beta \frac{F_N}{H} v \tag{16}$$

where A_a is the apparent contact area of spool and valve sleeve. Substitute Eq. (16) into Eq. (15)

$$D = \frac{C}{A_a} \cdot \frac{dV}{dt} \tag{17}$$

And Eq. (13) can be changed into

$$\frac{d\sigma_s}{dt} = -\frac{2C}{A_a} \cdot \frac{dV}{dt} \cdot \sigma_s^{-1}\left[\int_{-\infty}^{+\infty} z \cdot (z-h)^{\gamma} \cdot \varphi(z,t)dz + \int_{-\infty}^{+\infty} z \cdot \varphi(z,t)dz \cdot \int_{-\infty}^{+\infty} (z-h)^{\gamma} \cdot \varphi(z,t)dz\right] \tag{18}$$

If we take γ = 1 which means the asperity wear velocity and the embed depth of asperity follow a linear relationship. Then we can get a simplification of Eq. (13)

$$\frac{d\sigma_s}{dt} = -\frac{2C}{A_a} \cdot \frac{dV}{dt} \cdot \sigma_s^{-1} \cdot \{E[Z^2(t)] + \{E[Z(t)]\}^2 - 2hE[Z(t)]\} \approx b_1(\sigma_s - b_2)\frac{dV}{dt} \tag{19}$$

where the coefficient $b_1 < 0$ and $b_2 > 0$.

3.4 Wear Coefficient K

In molecular mechanical friction theory, friction is considered the process of overcoming mechanical engagement of asperities and molecular attraction, thus friction force is the sum of mechanical effect resistance and molecular effect resistance which takes the form according to reference [7]

$$F_\mu = F_{fi}A_f + F_{ji}A_j \tag{20}$$

where A_f and A_j are areas of molecular effect and mechanical effect of the real metal-to-metal contact areas between spool and valve sleeve respectively, F_{fi} and F_{ji} are friction force per unit real contact area caused by molecular effect and mechanical effect respectively and they can be calculated by the following equations

$$F_{fi} = S_f + B_f p_r \tag{21}$$

$$F_{ji} = B_j p_r^2 \tag{22}$$

where S_f is the tangential resistance of molecular effect which is related to the surface cleanliness of friction pair, B_f is the surface roughness effect coefficient, B_j is the normal load effect coefficient, p_r is the normal load per real metal-to-metal contact area of the friction pair. Substitute Eqs. (21) and (22) into Eq. (20), we have

$$F_\mu = A_f(S_f + B_f p_r) + A_j B_j p_r^2 \tag{23}$$

The areas of mechanical and molecular effect of the friction pair follow the relationship

$$A_j = \gamma A_f \tag{24}$$

$$A_m = A_j + A_f \tag{25}$$

and

$$W_a = p_r A_m \tag{26}$$

where γ is the proportionality constant, W_a is the asperity load between spool and valve sleeve. Substitute Eqs. (24), (25) and (26) into Eq. (23) will lead to

$$F_\mu = \frac{S_f}{1+\gamma} A_m + \frac{B_f}{1+\gamma} W_a + \frac{\gamma B_j}{1+\gamma} W_a p_r \tag{27}$$

The friction coefficient can be obtained by coulomb's law

$$f = \frac{F_\mu}{W_a} = \frac{S_f}{1+\gamma} \cdot \frac{A_m}{W_a} + \frac{B_f}{1+\gamma} + \frac{\gamma B_j}{1+\gamma} \cdot \frac{W_a}{A_m} = \frac{\tau_0}{p_r} + \kappa + \alpha p_r \tag{28}$$

where $\frac{\tau_0}{p_r} + \kappa$ as a whole is the molecular component of friction coefficient f among which κ is the molecular bond strength coefficient, αp_r is the mechanical component of friction coefficient f.

According to reference [9], asperity load W_a and real contact area A_m can be calculated by the following two equations

$$W_a = \frac{4}{3} n A_a E' R^{1/2} \sigma_s^{\frac{3}{2}} \left[F_{\frac{3}{2}}\left(\frac{h}{\sigma_s}\right) - F_{\frac{3}{2}}\left(\frac{h}{\sigma_s} + \frac{\delta_p}{\sigma_s}\right) \right] + \frac{2}{3} \pi n A_a H R \sigma_s \left[F_1\left(\frac{h}{\sigma_s} + \frac{\delta_p}{\sigma_s}\right) \right] \tag{29}$$

$$A_m = n \sigma_s R \pi A_a \left[F_1\left(\frac{h}{\sigma_s}\right) + F_1\left(\frac{h}{\sigma_s} + \frac{\delta_p}{\sigma_s}\right) \right] \tag{30}$$

where

$$F_v(u) = \frac{1}{\sqrt{2\pi}} \int_u^{+\infty} (t-u)^v e^{\frac{t^2}{2}} dt \tag{31}$$

So

$$p_r = \frac{W_a}{A_m} = \frac{4\pi E' [F_{\frac{3}{2}}(\frac{h}{\sigma_s}) - F_{\frac{3}{2}}(\frac{h}{\sigma_s} + \frac{\delta_p}{\sigma_s})]}{3R^{\frac{1}{2}} [F_1(\frac{h}{\sigma_s}) + F_1(\frac{h}{\sigma_s} + \frac{\delta_p}{\sigma_s})]} \sigma_s^{\frac{1}{2}} + \frac{2H[F_1(\frac{h}{\sigma_s} + \frac{\delta_p}{\sigma_s})]}{3[F_1(\frac{h}{\sigma_s}) + F_1(\frac{h}{\sigma_s} + \frac{\delta_p}{\sigma_s})]} \tag{32}$$

Substitute Eq. (32) into Eq. (28), we can get the relationship between friction coefficient f and Rms of composite profile σ_s as follows

$$f = \frac{\tau_0}{p_r} + \kappa + \alpha p_r = \frac{1}{A_1 \sigma_s^{\frac{1}{2}} + A_2} + A_3 + A_4 \sigma_s^{\frac{1}{2}} \tag{33}$$

where A_1, A_2, A_3 and A_4 are positive coefficients.

Combining the empirical relationship between wear coefficient K and friction coefficient f

$$\lg K = 5 \lg f - 2.27 \tag{34}$$

we can get the following equation

$$K = \left(\frac{1}{a_1 \sigma_s^{\frac{1}{2}} + a_2} + a_3 + a_4 \sigma_s^{\frac{1}{2}} \right)^5 \tag{35}$$

where a_1, a_2, a_3 and a_4 are positive coefficients which are related to the physical and chemical properties and lubrication condition of spool and valve sleeve.

3.5 Dynamic Wear Model

In summary, the dynamic wear model of aero-hydraulic spool valve established in this paper with wear volume V and Rms of composite profile σ_s as the two independent state variable is a first order double variable dynamic system model

$$\frac{dV}{dt} = K\beta \frac{W_a}{H} v$$

$$\frac{d\sigma_s}{dt} = b_1(\sigma_s - b_2)\frac{dV}{dt}$$

$$K = \left(\frac{1}{a_1 \sigma_s^{\frac{1}{2}} + a_2} + a_3 + a_4 \sigma_s^{\frac{1}{2}} \right)^5$$

$$\beta = 1 - \exp\left\{ -\left[\frac{1.4 \times 10^{-8} V_m^{\frac{1}{3}}}{v t_0} \right] \exp\left(-\frac{E_c}{RT} \right) \right\}$$

where a_1, a_2, a_3, a_4, b_1 and b_2 are parameters which can be estimated by fitting the data from wear test on a aero-hydraulic spool valve where the wear volume and surface roughness of spool and valve sleeve under different oil temperature T and relative motion velocity v at different time t is observed.

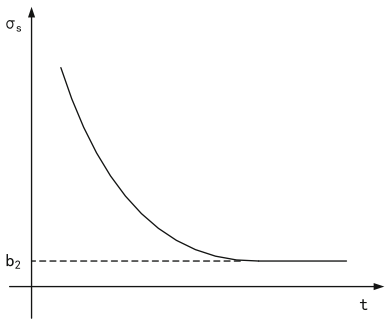


Fig. 8 Variation of surface roughness during wear

4 Analysis and Discussion

4.1 Variation of Surface Roughness During Wear

By solving Eq. (19), we can get

$$\sigma_s = b_3 e^{b_1 \int_0^t \frac{dV}{dt} dt} + b_2 \tag{36}$$

where $b_1 < 0$, $b_2 > 0$, $b_3 > 0$. Form Eq. (36) we can see that the variation of surface roughness during wear caused by the asperity truncating process follows the form of an exponential model where the trend of the change of σ_s with time can be estimated by the symbol of the three coefficients listed above. Rms of composite profile σ_s or surface roughness will decrease with time and finally reaches a stable value of b_2 as shown in Fig. 8.

4.2 Variation of Wear Rate During Wear

According to Eq. (35), let $K' = K^{\frac{1}{5}}$ will lead to

$$K' = \frac{1}{a_1 \sigma_s^{\frac{1}{2}} + a_2} + a_3 + a_4 \sigma_s^{\frac{1}{2}} \tag{36}$$

while K and K′ will exhibit the same trend of variation with σ_s, and

$$\frac{dK'}{d\sigma_s} = \frac{1}{2}\left[a_4 - \frac{a_1}{(a_1 \sigma_s^{\frac{1}{2}} + a_2)^2}\right] \sigma_s^{-\frac{1}{2}} \tag{37}$$

When $\sigma_s = \sigma_{sj} = \frac{1}{\sqrt{a_1 a_4}} - \frac{a_2}{a_1}$, K′ and also K will reach the minimum as shown in Fig. 9.

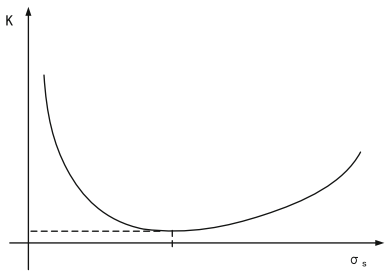


Fig. 9 Relationship between K and σ_s

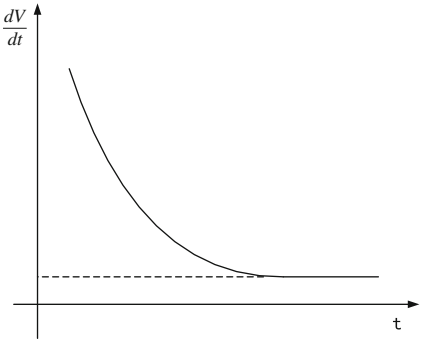


Fig. 10 Variation of wear rate during wear

The wear coefficient K is mainly affected by molecular force when $\sigma_s \leq \sigma_{sj}$ where the wear coefficient K will exhibit a sharp increase as σ_s decreases, which means a very smooth surface will lead to a very large wear coefficient. The wear coefficient K is mainly affected by mechanical force when $\sigma_s > \sigma_{sj}$ where the wear coefficient K will increase as σ_s increases which means a very rough surface will also lead to a large wear coefficient. It can be theoretically deduced that $\sigma_{sj} = \frac{1}{\sqrt{a_1 a_4}} - \frac{a_2}{a_1}$ is a very small value which can't be reached by the general machining process to achieve such a high quality surface within the scope of molecular force. So in the case of aero-hydraulic spool valve, the wear coefficient K is an increasing function of σ_s.

In combination with Eq. (16), wear rate $\frac{dV}{dt}$ is proportional to wear coefficient K which gives the variation of wear rate with time during wear process as shown in Fig. 10.

It can be seen from Fig. 10 that the wear rate of aero-hydraulic spool valve will decrease with time and finally reach a stable value which corresponds to the running-in and stable wear period. As can be seen from Fig. 9 that the surface roughness decreases with time which gives a negative feedback to the wear rate, so the spool valve will not enter the severe wear period in its life time under normal working condition.

References

1. Yan J, Wenhan W, Zhanliang J (2011) Analysis and improvement on the clamping issue of high precision hydraulic spool valve. Mach Tool Hydraulics 39:112–115 (in Chinese)
2. Anlin W, Yaning D, Pengjv Z (2011) Robust design on the clamping issue of hydraulic valve. J Shanghai Jiaotong Univ 45:1637–1652 (in Chinese)
3. Liang W, Shudong Y (2012) Wear life prediction for a large diameter spool valve. Mach Des Manuf 8:234–236 (in Chinese)
4. Hu YZ, Li N, Tonder K (1991) A dynamic system model for lubricated sliding wear and running-in. J Tribol 113:499–505
5. Archard JF (1953) Contact and rubbing of flat surfaces. J Appl Phys 24(981):981–988
6. Stolarski TA (1996) A system for wear prediction in lubricated sliding contacts. Lubr Sci 8:315–351
7. Shizhu W, Ping H (2011) Principles of tribology, 3rd edn. Tsinghua University Press, Beijing (in Chinese)
8. Golden JM (1976) The evolution of asperity height distribution of a surface subjected to wear. Wear 39:25–44
9. Chenbo Ma, Hua Z (2008) Theoretical analysis on the asperity load of the cylinder-piston ring during run-in phase. Veh Engine 4:32–35 (in Chinese)

Visualization Workflow Modeling System Research and Development Based on Silverlight

Wang Lei and Yuan Hongjie

Abstract Workflow technology based on computer application environment is one of the production and operations management techniques that currently rapidly developed. A Silverlight-based technology workflow editor design method has been proposed, highlighting its features of visual graphical user interface contrast poor web-based graphical workflow modeling techniques and weak performance defects. Referring Workflow Management Coalition standards and workflow process definition model, we built the workflow management system architecture applied the above workflow editor, and then briefly described development strategies of client application interface, process manager, workflow engine and data storage, and their design interface methods according to XML and WCF technology.

1 Introduction

With the development of industrialization and informationization, product development and project implementation has gradually shifted from completing by individual companies to a number of different agencies to promote concurrently. Product and project information management system software platform has also been implemented gradually from C/S transition to B/S architecture. Institutions are organized as dynamic alliance, their collaborative relationships are born with the birth of the project, and demised with completion of target. It is often necessary to reconstruct their business processes to accommodate the new projects to be launched. Therefore, the stiff structure that encoding process run into the application system is clearly not suitable to the new organizational model. There is a must for

W. Lei (✉) · Y. Hongjie
Beijing University of Aeronautics and Astronautics, Beijing 100191, China
e-mail: wanglei_0905@126.com

Y. Hongjie
e-mail: yuanhongjie@buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_123

Product and operations management systems to have a visualization workflow Designer compatible to B/S architecture.

Workflow is a computerized business process. Workflow management system that completes the workflow definition and management and self-propels workflow instance execution in accordance with predefined workflow logic is the software environment of its implementation. Real-world business process is abstracted to workflow model that uses a formal, machine-processable way to represent and includes all process information can be performed the service by the workflow software service system. The process information includes the beginning and end conditions, activities and navigation rules between the activities, applications, reference relationship between machines, and data definitions.

In this chapter, a Silverlight-based technology workflow editor design method has been proposed according to the research status of the workflow model, highlighting its visual graphical user interface features, and then briefly described development strategies of client application interface, process manager, workflow engine and data storage, and their design interface methods according to XML and WCF technology.

2 Theoretical Basis and Technical Presentations

2.1 Silverlight Architecture and Technical Content

Microsoft Silverlight has the ability to create interactive multimedia effects with rich next generation of Web applications (Rich Interactive Application, RIA) is a cross-platform browser technology of .NET Framework. This is because Silverlight unified server, web, desktop computer, managed code, dynamic languages, declarations and traditional program code as well as some Windows Presentation Foundation (WPF) feature. Although the Silverlight plug-in (Plug-In) is a client operating environment, its size does not exceed 5 MB, Silverlight development platform integrates numerous features and sophisticated technology, which makes it easy for developers to start to play its RIA features. As can be seen in Fig. 1, Silverlight development platform architecture consists of two main parts, and browser plug-in, see Table 1 for instructions.

Figure 2 is a Silverlight development framework, Silverlight Integrated many technologies into a single development platform, so that we can choose the appropriate development tools and programming language. The following is the content of Silverlight technology related to this article.

- WPF&XAML

Silverlight can be said that a subset of WPF. These techniques is sufficient to significantly expand XAML browser use XML-based declarative syntax to create a WPF project in order to create a more dazzling user interface, and to achieve a

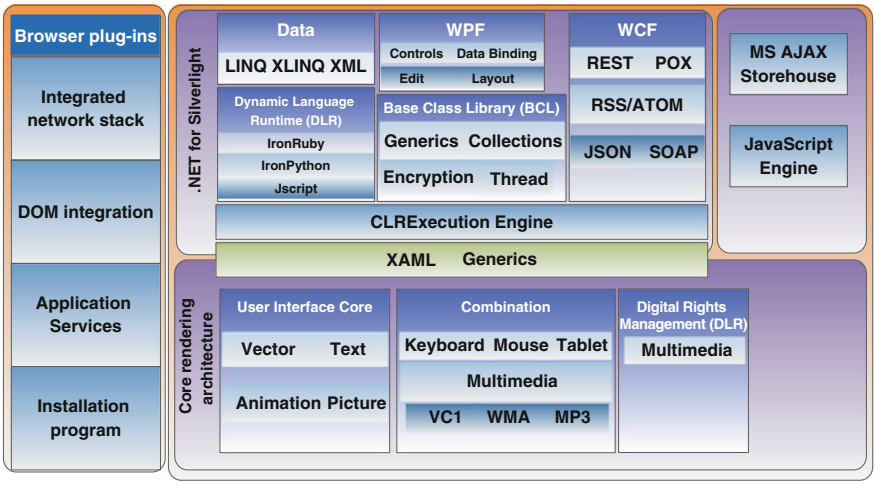


Fig. 1 Silverlight development platform architecture

Table 1 Description of framework development platform

Component	Explanation
Core rendering architecture	Provide the user interface (UI) and user interaction-oriented components and services, including user input, web application design specifically for lightweight UI controls, multimedia playback, digital rights management (DRM), data binding set and so on. The function of present aspects includes vector graphics, text, animation and pictures. Most importantly, these components, services, or features are specified their location in the layout by Extensible Application Markup Language (XAML)
.NET Framework for Silverlight	.NET Framework subset that contains components and libraries, including data integration, extensible Windows controls, networking, base class libraries (BCL), Garbage Collection and Common Language Runtime (CLR). Some .Net Framework for Silverlight component with application being deployed out, call these "Silverlight link library", because they are Component, and did not contain the Silverlight runtime (i.e. Silverlight plug-in). For example XLINQ, RSS/ATOM, XML serialization and DLR. These are designed for the launch of the new Silverlight UI controls
Browser plug-ins	Provide installation and updated features to simplify the first execution of the application install Silverlight plug-in processes for user, and provide automatic updating mechanism

deeper level of graphics, animation, multimedia, and other super-rich client functionality. Such as defining a graphic image, the user interface, operational behavior, animation and so on. Expanded to the browser-based user interface, so it presents a gorgeous effect far from the traditional HTML pages can match. That is, either a

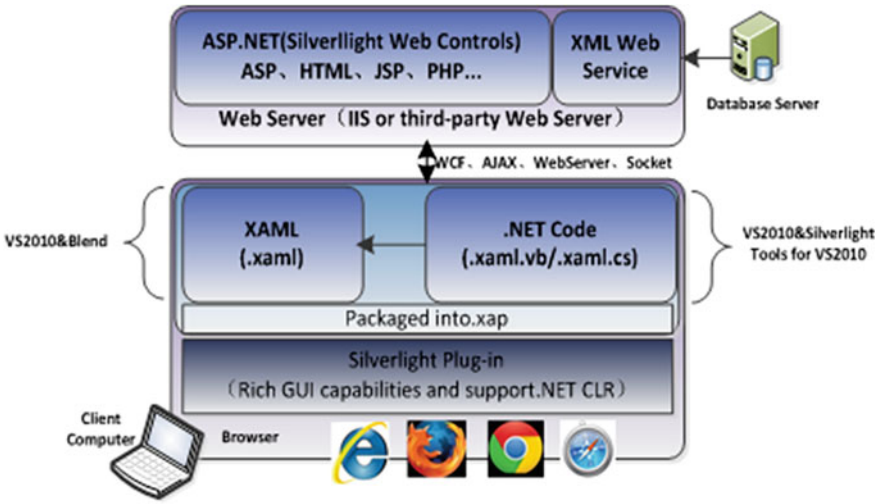


Fig. 2 Silverlight development framework

declarative XML markup to create visual user interface, defined conversion or animation effects and other dazzling appearance.

- Support Network

Silverlight can access the WCF, SOAP, ASP.NET AJAX and other services in order to receive XML, JON or RSS data via HTTP or TCP communication protocol. One admirable feature is that user can only download the required data content (such as XAML, components or multimedia data like images), with emphasis on downloads is to use asynchronous mode, so users will not interrupt the current operation.

- LINQ

Silverlight embeds LINQ, which means you can use native syntax and strongly typed objects of .NET Framework language that is easy to understand to access data.

2.2 Workflow Process Definition Meta-Model

Workflow model an abstract description of real-world business processes is a formalized conversion tool of computer internal representation. Workflow model uses a simple and intuitive form to abstract complex processes, providing a reference framework for describing the world. In order to provide a common way to realize information exchange, interpretation and implementation between multiple

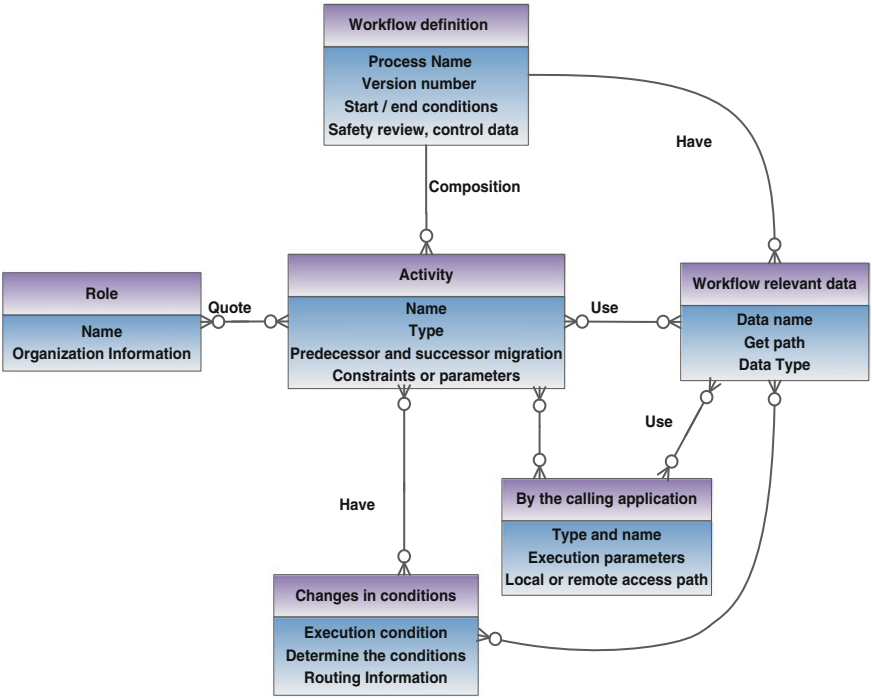


Fig. 3 Process definition metamodel

workflows, the Workflow Management Coalition to develop and define a process definition meta-model, including physical object and its properties element, such as workflow definitions, activities, roles, and changes of conditions, etc., the relationship shown in Fig. 3.

Workflow Management Coalition proposed workflow process definition language WPDL as a bridge between different workflow products. WPDL syntax similar to Backus—Naur form support data exchange between different workflow products. Various workflow definition model complete the data conversion and interaction by completing conversion between their own internal data representation and WPDL prescribed format .eXtensible Markup Language (XML) is a technology of cross-platform and depending on the content, and a powerful tool to processing structured document information. It is precisely because the XML characteristics of self-contained, self-describing as well as its powerful data storage and analysis capabilities, making it become a good WPDL carrier, XPDL workflow definition language came into being.

3 Silverlight-Based Visual Workflow Designer

3.1 Design Principles of Workflow Designer

Design and development methods of visual workflow designer this article describes is based on the workflow process definition and WFMC the relevant standards. As the designer runs on B/S architecture system, in order to solve user interface monotonous, poor modeling and cross-platform, heterogeneous and other issues on the practical application, used Silverlight combined with XML technology, enabling the user interface expressive force became strong and data conversion and processing more convenient.

Workflow Designer provides a graphical process definition, which includes three categories of objects:

Activities Block Diagram (triangle charts, pie charts) corresponding object that represents the workflow activity.

Rule arrowed lines (curves, polylines), this object represents a workflow transition.

Design panel design panel is a container object of the flow chart.

All these objects described as class have a user interface that designed by Silverlight user controls (xaml file in the user interface layout, write the background method in corresponding xaml.cs file). Activities and rule model design primarily depending on the use of the application for mouse events capture at public API framework FrameworkElement object. Attribute definition is relatively simple. Container object includes a collection of activities and rules and handle methods, and has a syntax and semantics checking capabilities to help modelers create the correct model, so that the system can express basic modeling business processes. Similarly, you can flexibly set according to different needs, increasing the role classes and applications to improve workflow process easy to understand and rigor, in the system with complex workflow model defines. According to the actual business process, user takes modeling interface element tools to create the appropriate workflow topology through a graphical workflow entities, and set workflow properties in the properties dialog box for the corresponding entity in full compliance with WfMC process definition specification. Workflow model convert into XPDL format through workflow model parser, including the definition of attributes, such as the number associated ID, name, position, application, relevant data fields, custom extended class, etc., All have been written in a structured form XPDL text to store, while parsed the XPDL text to various controls for Silverlight, showing a flow chart through XPDL parser.

3.2 Design Features of Visual Interface

Silverlight based on technique of running the C# code in the client's browser, we can put Silverlight understood as programming of C/S structure while achieve a the program of B/S structure, which makes the design classes accordance with above ideas to become a reality. This workflow design based on object-oriented concepts and design methods, such as class inheritance, time-triggered (Event Trigger), dynamic binding, messaging and multi-state, having the following characteristics: (1) Brush object is defined to support the achievement graphically display interface element object rendering, (2) By MouseLeftButtonUp event triggers and Canvas. Left&Contenter.Top property settings to achieve dragging to creating and modify the process, (3) By implementing Geometry realizing depending on different activities display different type of shapes, (4) Through the Line object and angle inheritance Canvas class settings to achieve the flexibility to move with the arrows and the migration of inflection, (5) By adding the Stack to achieve process model withdrawal and advance operation, (6) By introducing DispatcherTimer class object implements the mouse double-click and right-click event, (7) ComboBox control to achieve activities properties configuration, (8) Achieve Clone Interface adding method clone and delete, copy, realize objects copy and delete, (9) By adding a DoubleAnimation object implements the model fades, (10) By using the scroll bar slider drag achieve scaling process model. Designer Modeling interface shown in Fig. 4. These are just some features of this editor, designers can design different features according to related needs combining Sliverlight.

3.3 XPDL Model Transformation and Analysis

The system dynamically generates a topology with corresponding XPDL process text stored in memory the same time with visual process modeling. XPDL an XML-based process definition language describes the process (WorkFlow) structure, the

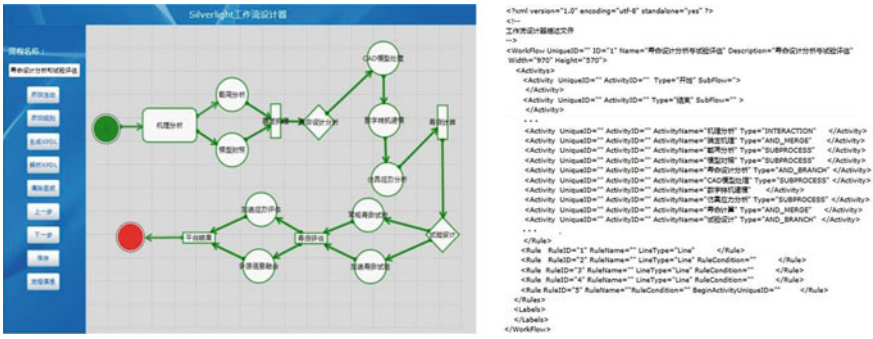


Fig. 4 View of client graphical and view of the client XPDL

```
var partNos = from item in xele.Descendants("Activity") select item;
foreach (XElement node in partNos)
{

    activityID = node.Attribute(XName.Get("ActivityID")).Value;
    activityName = node.Attribute(XName.Get("ActivityName")).Value;
    double.TryParse(node.Attribute(XName.Get("PositionX")).Value, out temd);
    activityPosition.X = temd;
    double.TryParse(node.Attribute(XName.Get("PositionY")).Value, out temd);
    activityPosition.Y = temd;
    int.TryParse(node.Attribute(XName.Get("ZIndex")).Value, out zIndex);

    Activity a = new Activity(this);
    a.Position = activityPosition;
    a.ActivityID = activityID;
    a.ActivityName = activityName;
    a.ZIndex = zIndex;
    AddActivity(a);
}
```

Fig. 5 XPDL parsing code examples

definition of which activities (Activity) and rules (Rule) properties through different markup syntax tree. Part of XPDL view of the model shown in Fig. 4.

Workflow Designer not only put the graphical view into XPDL text, but also parse XPDL text generating visual models in the designer. XPDL analytical method used in this chapter is LINQ to XML programming techniques .NET Framework provides, using multiple LINQ to XML classes which are XML programming interface in System.Xml.Linq namespace. LINQ model is more lightweight than the Document Object Model (DOM), then more convenient to use. Using LINQ to XML can load XPDL from a variety of data sources, such as a string, XmlReader, TextReader or file. The most important advantage of LINQ to XML is its integration with Language-Integrated Query (LINQ). Because achieve this integration, so we can write a query an XML document in memory to retrieve the workflow elements and the properties collections. By making the query result as constructor parameters of XElement and XAttribute object to achieve a powerful method for creating XML trees. This method is called “Constructor”, using this method, developers can easily convert XPDL tree view to a graphic object, and then generate visual workflow model. Using LINQ to XML traversal XPDL view in XPDL tree is relatively simple use methods. Elements and Element method provided by XElement and XAttribute class provides the way of targeting to one or some elements. Part of the parser, see Fig. 5.

4 Design of Visual Workflow Management System

4.1 Architecture of Workflow Management System

The preceding gives design method of visual workflow designer based on Silverlight technology, in this section, we determine visual workflow management systems framework according to WfMC reference model, combined with .Net Framework platform. The entire system can be divided into five parts, interacting via respective interfaces between the five parts (Fig. 6). Process definition tools that defined business processes finally stored as XPDL data format based on business model is visual workflow designer. The workflow engine is responsible for reading XPDL data files and workflow analysis, and generate running controller of process instance and cure process-related objects. Management and monitoring makes administrators can deploy activities instances and terminate a process instance through the interface with the implementation of the workflow engine. Client applications, mainly for users log in under the appropriate permissions to access specific task table. Program calls can be linked to access different systems and procedures. The architecture in addition to the uses XPDL Schema to realize Workflow component interface model converse, also uses XML serialization/deserialization and WCF data communications technology.

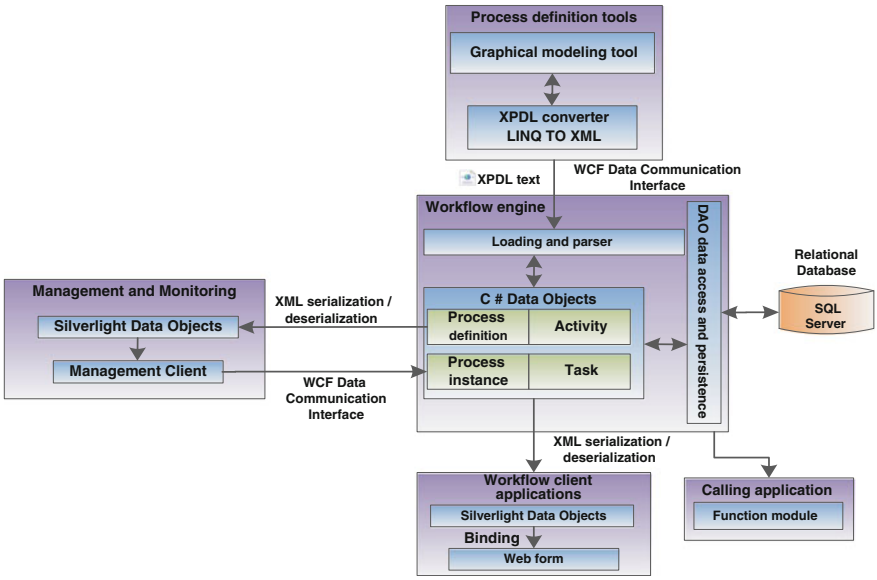


Fig. 6 Workflow management system architecture

4.2 Key Technology of Interface

WCF is a collection of a variety of communication methods, and data communications application development interfaces of service-oriented architecture with advantages of unity, interoperability, security, trustworthiness and security, along with a simple programming model including service contracts, the host and port. Use contract to define protocol between communication, unify communications protocol by the way of bindings (Fig. 7). Clients and servers need to be set for WCF programming, especially for server configuration document preparation is very important, many of the major configuration items in this file <system.serviceModel> one of basic set elements on WCF, configure the sample program shown in Fig. 8. Various workflow objects have different forms can be achieved interaction between client and server by using WCF services.

Serialization is a process of converting the object into the form easy to transmission. Objects can be serialized, using HTTP to transmit over the Internet between client and server. On the other hand, deserialization is to reconstruct the object in the stream. The most important method of XmlSerializer class the core class of the XML Serialization is Serialize and Deserialize. XmlSerializer creates C# file and compile it to .dll files to perform this serialization. XmlSerializer class can further serialize an object and generate an encoded XML stream conforms "Simple Object Access Protocol" (SOAP). Data objects obtained by DAO can be transmitted via WCF, while relational data which is a formation of deserialized Silverlight data objects display to the user through data binding in the client data objects.

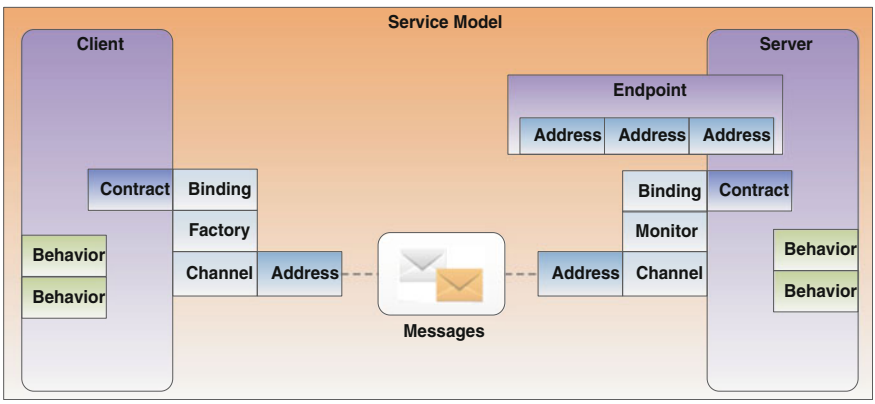


Fig. 7 WCF programming model

```
<?xml version="1.0" encoding="UTF-8"?>
- <system.serviceModel>
  - <behaviors>
    - <endpointBehaviors>
      - <behavior name="MyendpointBehavior">
          <dataContractSerializer maxItemsInObjectGraph="2147483647"/>
        </behavior>
      </endpointBehaviors>
    - <serviceBehaviors>
      - <behavior name="">
          <serviceMetadata httpGetEnabled="true"/>
          <serviceDebug includeExceptionDetailInFaults="false"/>
          <dataContractSerializer maxItemsInObjectGraph="2147483647"/>
        </behavior>
      </serviceBehaviors>
  </behaviors>
  - <bindings>
      <basicHttpBinding/>
    - <customBinding>
      - <binding name="customBinding0">
          <binaryMessageEncoding/>
          <httpTransport/>
        </binding>
      </customBinding>
  </bindings>
  <serviceHostingEnvironment multipleSiteBindingsEnabled="true"/>
  - <services>
    - <service name="WorkFlow">
        <endpoint contract="WorkFlow" bindingConfiguration="WorkFlow" binding="customBinding" address=""/>
        <endpoint contract="IMetadataExchange" binding="mexHttpBinding" address="mex"/>
      </service>
  </services>
</system.serviceModel>
```

Fig. 8 Workflow service configuration structure

5 Conclusion

With the increasing development of workflow management system, higher requirements about visualization modeling tools and transferring and sharing data between systems have been proposed. Graphical modeling approach is preferred as the user interface, while developing workflow model designer on Silverlight whose characteristics of rich interactive (RIA) are ideal for GUI development is a very good choice. Meanwhile, it also meets requirements of the workflow system for heterogeneous data transfer and sharing combined with WCF services and XML technology. In this chapter, we research and design a workflow modeling system based on Silverlight. This system which applied to the project management system of software platform has a good effect. It is confident to say that the technology route has certain reference value and positive role for application and promotion of B/S architecture workflow modeling technology.

References

1. Wang J (2007) Web-based visual workflow modeling and analysis. J Tongji Univ (ShangHai)
2. Zhang L (2009) Silverlight2.0 development technology Pristine. Science Press, Beijing
3. Hou Z, Yu Z, Feng Z (2010) Record of workflow management systems development. China Railway Press, Beijing

4. Ma D (2004) Visualization workflow management system design and development. Manuf Autom (Beijing)
5. Wang K (2005) Application of JGraph in RRFlo workflow model designer. Manuf Autom (Beijing)
6. Meng Y, Wang F, Zhang R (2009) Design of workflow engine based on WCF. In: World congress on software engineering

Application of Simulation Method in the Structural Failure Analysis of an Airborne Product

Demiao Yu, Zhilqiang Li and Shimin Zhai

Abstract The reasons of some structural failures cannot be defined. In the reliability enhancement test of an airborne product, a simulation method has been proposed to calculate fatigue damage of random vibration power density spectrum of the fault point within frequency domain. The exact failure times can be calculated accurately and the most possible failure mechanism can be analyzed. In Sect. 5 (Analysis and Verification), the correctness of the results of the simulation analysis has been verified and an optimization design of the product structure has been proposed.

Keywords Simulation method · Fatigue damage · Frequency domain

1 Introduction

Some of product faults which are exposed in the reliability enhancement test of electromechanical products are electronic device faults, others are mechanical structure faults. Electronic device faults are normally found in the output data during testing. While the mechanical structure faults usually cannot be discovered until series of serious subsidiary faults happened, such as the poor contact caused by fastener breaks, the short circuit caused by metal parts off and so on. These are always occurred after abnormal conditions. In this situation, if the faults cannot be found and repaired immediately, it would cause unnecessary damages to equipment

D. Yu (✉) · Z. Li
School of Reliability and Systems Engineering, Beihang University, Beijing, China
e-mail: yudemiao1987@sina.com

Z. Li
e-mail: leezhq@buaa.edu.cn

S. Zhai
School of Mechanical Engineering and Automation, Beihang University, Beijing, China
e-mail: aazhsm1987@126.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_124

which would increase the testing time and cost. In this chapter, the finite element techniques are used to do dynamic analysis for products and the fatigue calculating has been conducted. The chapter compared the correctness of FEA analysis with test report, and accurately located the failure time of the structural part. At the last Corroborated the experimental results which cannot be explained proposed optimization design and verified the result.

2 The Implementation of Reliability Enhancement Test

The research and application of the Reliability Enhancement Test started since 1960s and the test did not become mature until 1990s. The test is now widely used to largely enhance the reliability of aerospace products. Reliability Enhancement Test is kind of destructive test. In order to ensure the continuity of the experiment and protect the products, a scientific testing program needs to be developed. Because the designing of testing program is not the focus of the chapter, there is only a brief description of the process in Fig. 1.

2.1 Product Description

The product in this test and simulation is Cockpit Voice Recorder (CVR). Its main function is to receive flight data, voice, video information of Data Acquisition Unit and complete the coding. After that the data will be partitioned recorded in protective solid state memory. The environment of the CVR is complex. To be able to record data in any extreme environment including vibration, shock, high and low temperature, explosion shock is a basic requirement of CVR. Therefore, improving the reliability of CVR is very important.

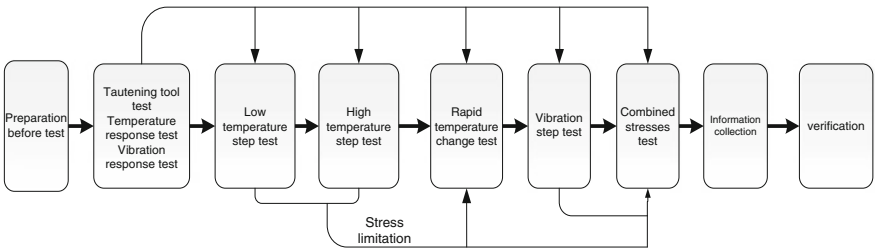


Fig. 1 Process of reliability enhancement test

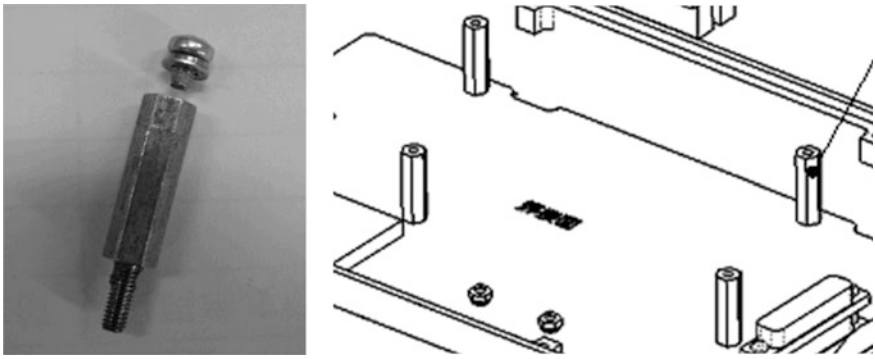


Fig. 2 The broken screw and product structure

2.2 Fault Information of Reliability Enhancement Test

There is only one fault occurred throughout the test and it happened in the vibration step test. Ethernet disconnected when vibration stepping to 20 G. After the investigation of the product, found one of the six screws in power board broken in the connection. The broken screw and product structure are shown in Fig. 2. The specification of screw is M2.5Cr18Ni9Ti [1].

2.3 Failure Analysis

The product uses six pillars to fix the CPU board on the chassis and uses another six pillars to fix power board on CPU board This makes the power board hanging instead of fixed on the side, which leads to the screw fatigue fracture as the screw stresses too much. The next section will use simulation method to speculate mode of failure and calculate screw life with frequency domain analysis method.

3 Calculation of Vibration Fatigue Life Within Frequency Domain

Frequency domain analysis methods are often used to predict fatigue life with random load. The general way is to convert the Power Spectral Density Function (PSD) to stress amplitude Probability Density Function (PDF). After years of development, researchers have developed a variety of models. Generally, different ways of calculating PSD would result different models. The i-th spectral moment is defined as:

$$m_i = \int_0^{\infty} f^i G(f) df \tag{1}$$

From Ref. [2], most fatigue life prediction models have their own scope of application. In order to maximize the accuracy of life calculation, different models are used depending on the spectral width coefficients. The spectral width coefficient is defined as:

$$\varepsilon = \sqrt{1 - \frac{m_2^2}{m_0 m_4}} \tag{2}$$

This chapter calculated the spectral width coefficient of the related PSD response of simulation. Dirlik method is to be used to calculate vibration fatigue life when spectral width coefficient is greater than 0.5. This method is more accurately than other methods regarding to describe the stress amplitude distribution within stress amplitude range when spectral width coefficient is bigger than 0.5. The formula of Dirlik empirical probability density function is defined as:

$$\mathrm{P(S)} = \frac{\frac{\mathrm{D_1}}{\mathrm{Q}} \mathrm{e}^{\frac{-\mathrm{Z}}{\mathrm{Q}}} + \frac{\mathrm{ZD_2}}{\mathrm{R}^2} \mathrm{e}^{\frac{-\mathrm{Z}^2}{2\mathrm{R}^2}} + \mathrm{D_3 Z e}^{-\frac{\mathrm{Z}^2}{2}}}{2\sqrt{\mathrm{m_0}}} \tag{3}$$

$$\mathrm{R} = \frac{e - x_m - D_1^2}{1 - e - D_1 + D_1^2};\ x_m = \frac{m_1}{m_0}\sqrt{\frac{m_2}{m_4}};\ D_1 = \frac{2(x_m - e^2)}{1 + e^2};\ D_2 = \frac{1 - e - D_1 + D_1^2}{1 - R};$$

$$D_3 = 1 - D_1 - D_2;\ \mathrm{Q} = \frac{1.25(e - D_3 - D_2 R)}{D_1};\ \mathrm{Z} = \frac{S}{2\sqrt{m_0}};\ \mathrm{E(P)} = \sqrt{\frac{m_4}{m_2}};$$

$$\mathrm{e} = m_2/\sqrt{m_0 m_4}$$

"e" is the irregular factor of spectral type; S is the stress change process [3, 4]. After calculating the amplitude probability density, this chapter used Miner linear cumulative damage theory to calculate the life [5]. The formula of fatigue damage is shown as:

$$\mathrm{D} = \sum D_i = \sum \frac{n_i}{N_i} = \sum \frac{E(P) T_i P(S_i) \Delta S}{k S^{-\mathrm{b}}} = \frac{E(P)}{k} T \int_0^{+\infty} P(S) S^b dS \tag{4}$$

In the formula 4, n_i represents the number of cycles on the stress level of S_i, N_i represents the fatigue life on the stress level of S_i. k and b are material constants of SN curve. T_i is the cycle time within the stress level. T represents the life time. The formula of SN curve is shown as:

$\mathrm{N} = \mathrm{k}S^{-\mathrm{b}}$ The time to failure of target object can be obtained by formula 2.

4 Analysis of Simulation Failure

The product of finite element model is finite element analysis software called ANSYS. The model is meshed via appropriate way and the components of the product are filled in with materials. In the virtual environment the same method is used to fix and constraint the model as enhancement test.

Comparing the simulation acceleration response of corresponding points with the response of measurement points in enhancement test, the frequency of modulus 1–3 of the product and the frequency of virtual product are the almost same. The modeling is accurately reflecting the true situation of product. Figure 3 is the comparison of simulation response with response of enhancement test while the vibration level is 18 G.

The random vibration spectrum is inputted into computer according to the value of vibration from enhancement test. The random vibration spectrum is listed in Fig. 4. Then get the break screw's equivalent stress and the PSD response spectrum. The levels of vibration and corresponded PSD input spectrum are shown in Table 1.

Using matlab to do Dirlik density model for the output of screw fracture PSD can calculates the accumulated fatigue damage under different stress conditions.

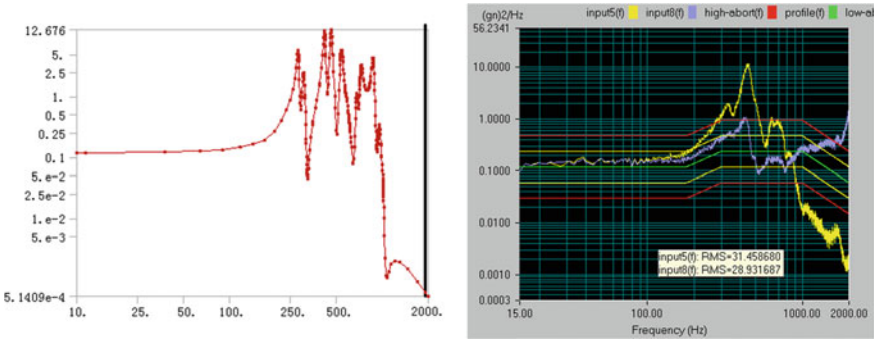


Fig. 3 The comparison of simulation response with response of enhancement test while the vibration level is 18 G

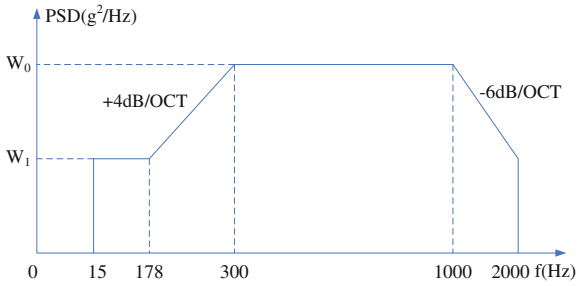


Fig. 4 Random vibration spectrum

Table 1 The levels of vibration and corresponded PSD input spectrum

Level/G	W_0 (g^2/Hz)	W_1 (g^2/Hz)	W_2 (g^2/Hz)
4	0.01166	0.00583	0.002915
6	0.02624	0.01312	0.00656
8	0.04664	0.02332	0.01166
10	0.07288	0.03644	0.01822
12	0.10494	0.05247	0.026235
14	0.14284	0.07142	0.03571
16	0.18656	0.09328	0.04664
18	0.2362	0.1181	0.05905

Table 2 The results after calculating cumulative fatigue

Level/G	m_0	t (s)	ε	D_i	∑Di
4	27.667	600	0.57657	0.0323	0.0323
6	62.097	600	0.57657	0.0586	0.0909
8	110.37	600	0.57657	0.0895	0.1804
10	172.47	600	0.57657	0.1241	0.3045
12	248.34	600	0.57657	0.1622	0.4667
14	338.03	600	0.57657	0.2033	0.6700
16	441.49	600	0.57657	0.2473	0.9173
18	560.46	600	0.57657	0.2947	1.2120

In the vibration step test, the time of each vibration step lasts 600 s. The results are shown in Table 2.

It can be inferred that the screw probably broken because of fatigue damage. As in the result of simulation test shown in Fig. 5, the maximum stress of the screws

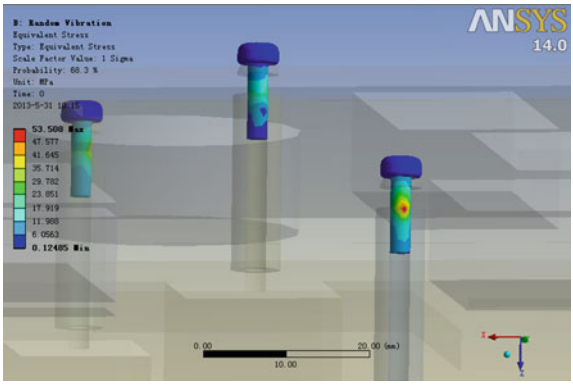


Fig. 5 The equivalent stress of screw while the level of vibration is 16 G

was within the limits of destruction. As shown in Table 2, the cumulative damage has reached 1.2 when the level of vibration stepped to 18 G. It can be deducted that the fatigue fracture has already occurred. In the enhancement test, the laboratory assistant did not find any significant faults on 18 G vibration level. But if continue to increasing the vibration level to 20 G, signal is interrupted. When the fault happened, the laboratory assistant did not found fault in the first time until the signal was interrupted on next vibration step. In this situation, it possible occurred unexpected even dangerous faults. According to the simulation results, the time can be calculated via Dirlik method. The time screw broken is 2–3 min after 18 G vibration level started and the failure mode is fatigue fracture.

5 Analyses and Verification

According to the report of enhancement test, the failure mode of screw is fatigue fracture. The reason why fatigue fracture occurred is the existence of uneven horizontal stress component due to speculation of fatigue striations direction [6, 7]. The scanning electron microscopy (SEM) analysis is shown as Fig. 6.

From macro perspective, fracture occurred at the root of the screw which supporting the weight of the power board. From the fracture morphology, the source of fracture is located in the surface of screw. The central fracture surface is apparently different from the border and the secondary cracks along the fatigue striation are showed obviously. Considering the direction of fatigue striation and the final position of the cracks, the screw's cracks is extended from the border to center along the vertical direction of fatigue striation and eventually break in the center. Thus, the results show the consistence of simulation analysis and the actual situation. It also provided a possible time of the fracture which helps further failure mechanism analysis.

According to the results of test, the following improvements and optimizations are adopted.

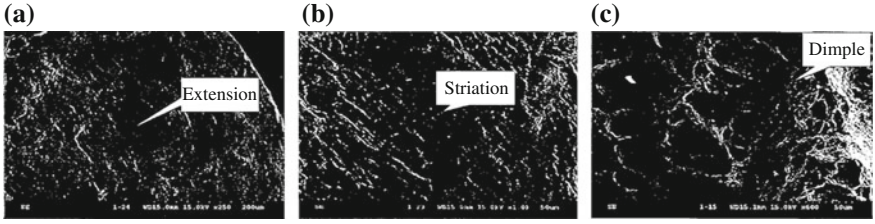


Fig. 6 SEM analysis results. **a** Extension of fatigue **b** Enlarge of fatigue striation **c** Morphology of dimple

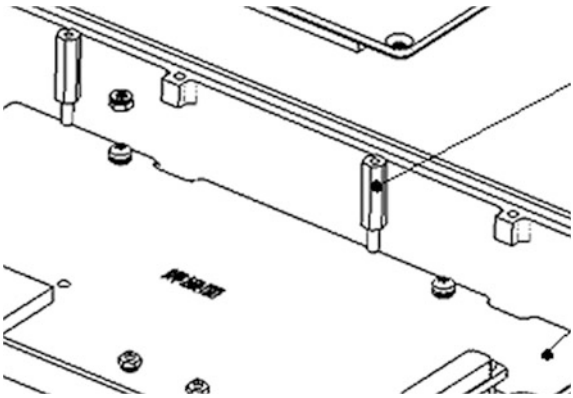


Fig. 7 Structures after improvement

(1) Changes of process

The screws are coated with Cable aldehyde phenolic glue which can effectively tighten the screws and prevent loosening before the screws are installed.

(2) Changes of structure

Structural changes are shown in Fig. 7. According to the results of simulation analysis of power board, the positions of screws and pillars have to be changed. The screws are combined directly to the base plate and in this way the installation of power board is improved. Vibration levels and horizontal stress will be reduced and the strength of the product in the horizontal direction is improved.

(3) Increasing of cushion

After improvements, the product successfully passes through the vibration stepping test and verification test which proofed that the changes are effective. The reliability of the product is improved.

6 Conclusions

Finite element simulation method is used to calculate fatigue life of an airborne product. It is determined that the time to failure of the screw is 2–3 min after the 18 G vibration begin. The most probable failure mode is determined by the stress information of simulation and supplements information, which cannot be explained in the enhancement test. The time of failure and failure mode is the strong evidence of pre-failure mechanism analysis and finally verified through SEM report. This chapter provided a new idea to calculate fatigue life of typical components which can reduce unnecessary testing costs and manpower. Based on simulation results, the improvements stated in Sect. 5 are optimized the product.

References

1. Liu M, Liu S (2011) Mechanical properties of metallic materials handbook. China Machine Press, Beijing
2. Yang W, Shi R (2011) Research on stress amplitude distribution of random vibration. Mach Des Res 27(6):16–20
3. Zhou M, Chen Z (2008) Fatigue life estimation and analysis of aircraft radar cover in random vibration environment. J Vib Eng 21:24–27
4. Sha Y, Zhang Z (2012) Fatigue life estimation method of random acoustic based on joint probability density. J Vib Meas Diagn 32:32–36
5. Peng L (2004) Calculation of structural fatigue life under broad band random loading spectrum. J Chang'an Univ (Nat Sci Edition) 24(1):76–78
6. Tan Y, Zhou Q (2007) Fracture analysis of screw. Heat Treat Metals 32:328–331
7. Zhao H, Wang J (2007) Fracture failure analysis of the high strength bolt. Heat Treat Metals 32:311–313

The Design and Implementation of the Non-electric Product Life Analysis and Calculation Software

Z. Li, Y. Chen and R. Kang

Abstract Non-electric products, undertaking power, transmission and other tasks, are important parts of the equipment. The normal operation of the non-electrical products is of great importance which leads to an urgent need of the quantitative analysis for life. Generally, the life of a product is predicted by tests, which cost a lot of time and money. Considering the lack of engineering life analysis methods, basic life models are collected through research of failure mechanism and survey on domestic and international data. The models are then classified as a library, based on which, a piece of life analysis and calculation software is designed and implemented. A life model library is built as well as the mapping relationships between life models and typical non-electric products and components by sorting the basic life models. And then a piece of software of life model library is designed. Furthermore, the implementation of the functions is comprehensively introduced from three aspects: project management module, data function module and database management module. Finally, an application of calculating the life of a typical non-electric product is illustrated to prove the software's practicality and efficiency in engineering. The result of the application can be used to make suggestions for design and improvement of the products.

Keywords Life calculation · Software design · Failure physics model

1 Introduction

Non-electric products, undertaking power, transmission and other tasks, are important parts of the equipment [1]. The normal operation of the non-electrical products is of great importance which leads to an urgent need of the life analysis.

Z. Li (✉) · Y. Chen · R. Kang
School of Reliability and Systems Engineering, Beihang University, Beijing, China
e-mail: lizi0309@163.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_125

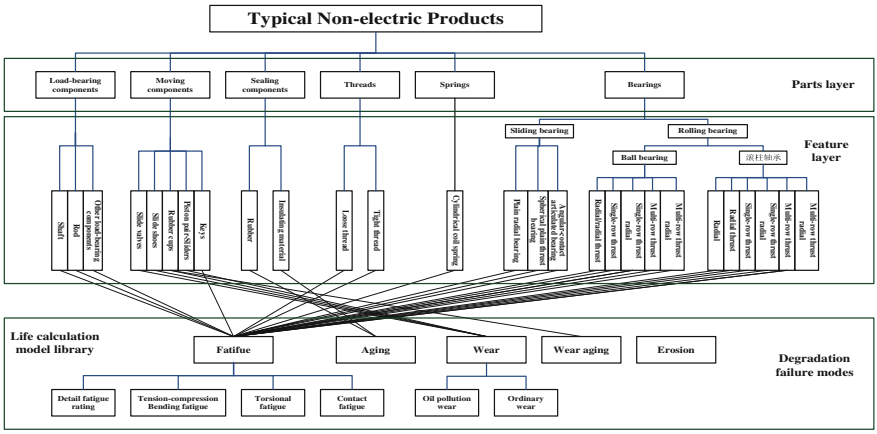


Fig. 1 The mapping relationships between the life calculation models and components

Considering the lack of engineering life analysis methods, basic life models are collected and classified as a life model library. Based on the library, a piece of life analysis and calculation software is designed and implemented.

2 Life Model Library

A comprehensive collection of life analysis and calculation models is made from all aspects such as domestic and foreign military standards, academic papers, monographs and teaching materials. A life model library is constructed and classified by failure modes such as fatigue, wear, aging, wear-aging and corrosion [2]. On the basis of the library, the mapping relationships between the life calculation models and components are researched according to the characteristics of different components Fig. 1.

3 Software Design

3.1 Function Design

The functions of the software are listed in Table 1.

3.2 Structural Design of Software

From the perspective of software architecture [3], life analysis software should establish interconnected modules in accordance with the procedure of life analysis

Table 1 Main functions of the software

Functions	Details
System management	System installation System register System login
Project management	Build a new project Modify existing projects Save a project Delete a project Print reports
Life calculation	Analyze and calculate different life indicators of products with different failure mechanisms and mission profiles
Database management	Project database management Model database management
Other functions	Software version information Software version information Help documentation

and calculation. Meanwhile, the software is supposed to have information management functions of products life analysis and calculation and set up the database for typical projects and life models. The overall architecture of the software is described in Fig. 2.

3.3 Design of Database

Database management system (DBMS) is the core tool for storage, analysis, statistics, evaluation, query and updating as well as an important part of the software. Basic functions of a DBMS are data manipulation, data logic operations, database structure manipulation, data retrieval and data report output, etc. Due to the small scale of the software database, Microsoft Access database is selected [4], using ADO as the connection object.

The designed E-R diagram of the project library is shown in Fig. 3. There are three data tables of the main entities involved in the project library, of which the model library is independent with only one table as shown in Table 2.

DBMS includes project database and component life model database, the former one saving structure information, components information and calculation results of projects, and the latter one containing typical examples of life calculations. Users can query, modify or delete existing projects from the project library, while the typical examples can be called directly from the model library.

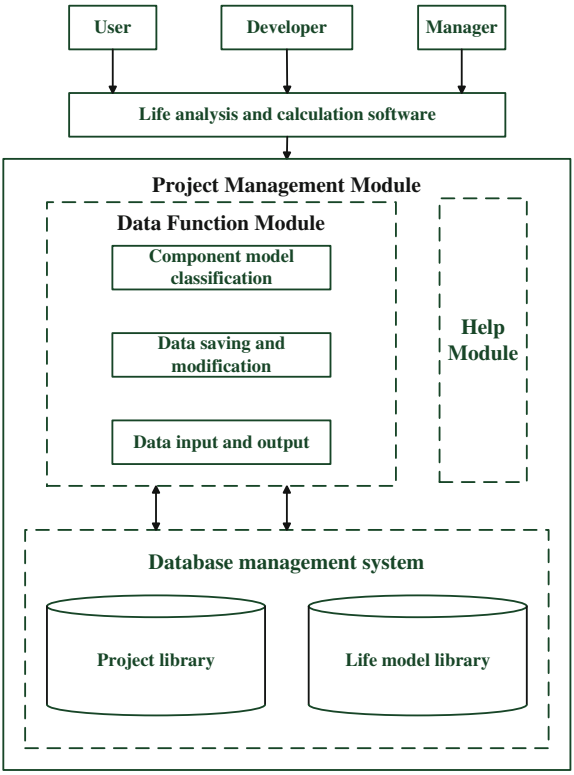


Fig. 2 Overall architecture of the life analysis and calculation software

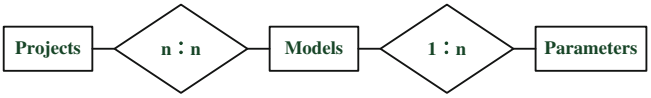


Fig. 3 E-R diagram of the project library

Table 2 Entity data tables

Table names	Fields
Project	Project ID, project name, creator, creation date, last updated date, project description, node number
Project structure	Project ID, node name, node description, node image, model no, form name
Project model information	Project ID, node name, control ID, control type, control name, parameter value, parameter type, parameter description
Model	Model ID, model no, control ID, control name, parameter value, parameter type, model description

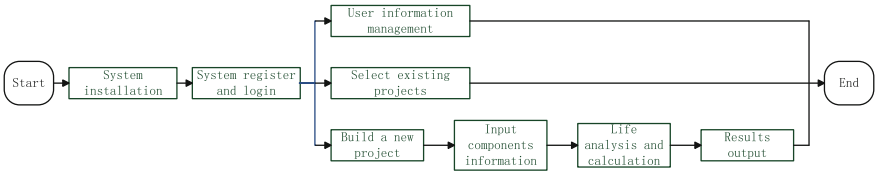


Fig. 4 Use flow chart

3.4 Software Data Flow

The use flow of life analysis software is shown in Fig. 4. The overall analysis process is illustrated in Fig. 5.

4 Software Implementation

The software is implemented through Microsoft Visual Basic with the modular thoughts [5].

4.1 Project Management Module

As the software framework, project management module covers the overall software in charge of project management, database management, version management and process management with CommonDialog and TreeView controls mainly [6]. This module helps realize building, opening, modifying, saving, saving as, report output of projects and model calculations as well as the updating of the database.

4.2 Data Function Module

The data function module, which calculates the life of the device, is the main part of the entire software. Text, OptionButton, ComboBox, PictureBox and MSFlexGrid controls are used in this module. Life indexes “Total Life” and “Time To First Overhaul (TTFO)” are alternative through the OptionButton control, which will be printed in the final report. Frames of different types of parameters are selectively displayed by ComboBox controls. Data-reference tables are recorded with Excel sheets, linked through PictureBox controls and displayed by MSFlexGrid controls.

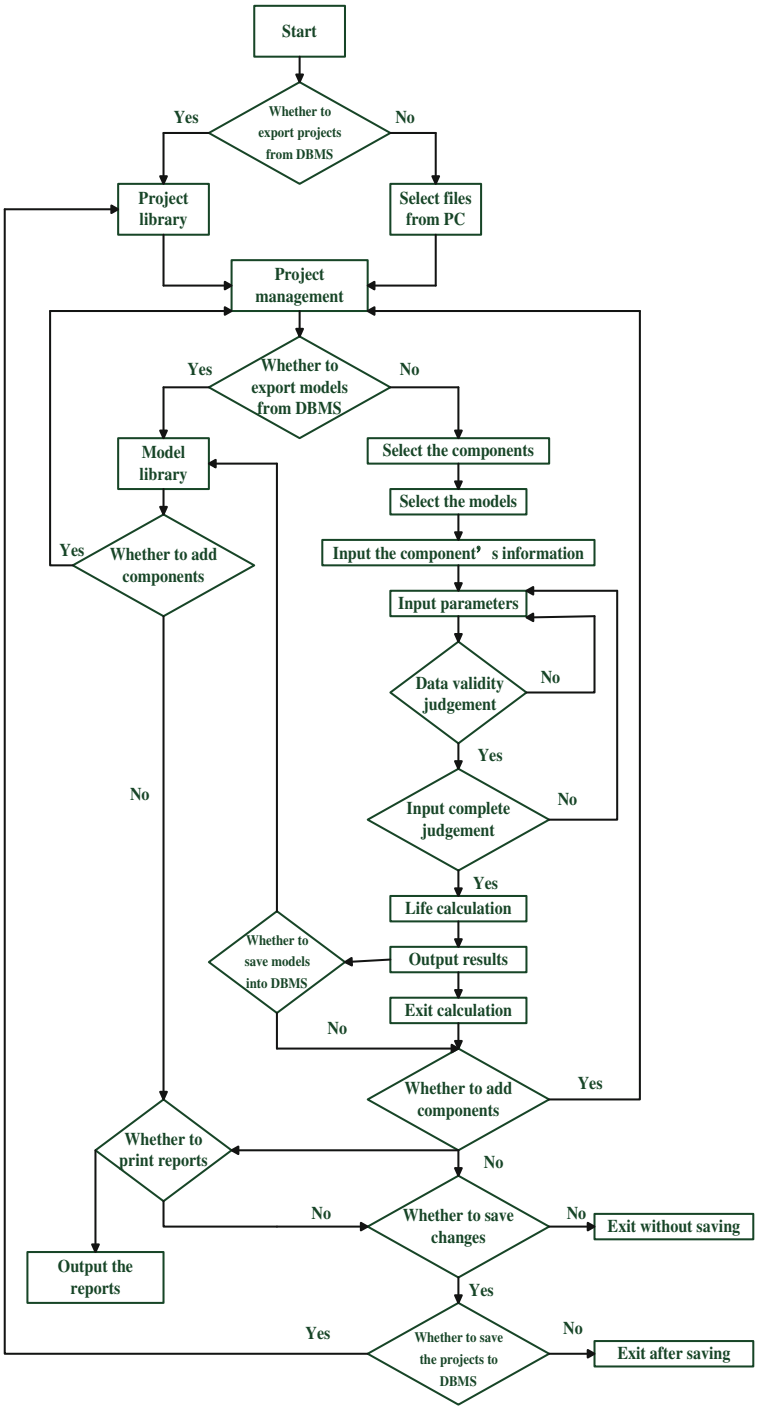


Fig. 5 Overall analysis process of the software

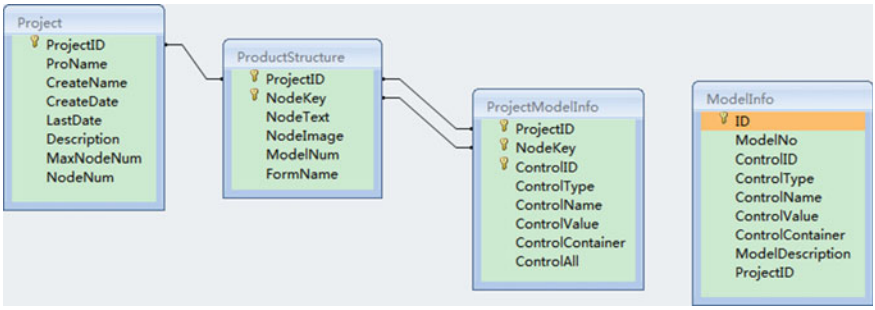


Fig. 6 Relationships in the database

4.3 Database Management Module

In the development of the software, basic project data and typical calculation examples are selected as the main targets of building a database by using CommonDialog controls and Access database with the DAO as the objects [7].

For the project library, project ID, creation date, creator, last updated date and calculation results are saved into the database. Each project has a unique ID as the main key. When queried, the above information will be shown through the MSFlexGrid control. On one hand, the on-going project can be imported into the database. On the other hand, projects in the database are able to be exported, shown in a tree structure by the TreeView control. Top of the tree is the project name, while the bottom is parameters. Models are classified by component name, failure mechanism model and other aspects in model library, independent of the projects. The database relationship is shown in Fig. 6. Projects and models are saved according to users' needs. All the data manipulations are realized by the SQL language.

5 Software Application

The software is used on an inner shaft subassembly, which belongs to a driven pump. The subassembly consists of an inner shaft, a tooth circle, an O-ring and a ball bearing.

The life models corresponding to each kind of failure mechanism of each component are selected through mechanism analyzing. Table 3 shows the analyzing and model selecting result.

Table 3 Mechanism analyzing and Model selecting result

Component	Failure mechanism	Failure model
Inner shaft	Involute spline contact fatigue	Involute spline contact fatigue life calculation model
Tooth circle	Adhesive wearing	Adhesive wear life calculation model
O-ring	Ageing	Rubber ageing life calculation model
Ball bearing	Contact fatigue	Ball bearing contact fatigue life calculation model

Table 4 Calculation result

Component	Life
Inner shaft	13742.56
Tooth circle	4100.8
O-ring	156792.9
Ball bearing	6326.6

After selecting models and inputting the parameters about structure, threshold, profiles and loads, Non-electric products' life can be analyzed and calculated through the software, and reported in documents. Calculation results are shown in Table 4.

The above results intuitively show the life of the components, which prove the practicality and efficiency of the software.

6 Conclusion

In this chapter, a non-electric product life analysis and calculation software is designed to systematically and conveniently calculate life indicators. Design and Implementation of the software is specified in the chapter. Finally, a driven pump is illustrated as the example to prove the validity and applicability of the software.

(1) A comprehensive collection of basic life models is made through the research. On the basis of the model library, a piece of life analysis and calculation software for non-electric products is developed with DBMS of projects and models.
(2) Modular thoughts and object-oriented design tools are used in the software which can calculate the life of non-electric products conveniently and accurately.
(3) Combined with DBMS, the program can easily invoke the projects and models saved in the database. In addition, the interface of the software is user-friendly together with the visualization of the data representation.

References

1. Xu H (2003) The theory of study of reliability design of mechanical parts. Dissertation, Xi'an University of Technology
2. Sun W, Chen Z, Wang C (2004) Mechanical failure analysis thoughts and failure cases. Trans Mater Heat Treat 25:1
3. Zhao T, Zeng S, Kang R et al (2000) Study on computer-aided reliability design and analysis system. Acta Aeronautica et Astronautica Sinica 21(3):206–209
4. Cai T (2011) Design and development of the assessing system based on access database in VB environment. Dissertation, Tianjin University
5. Sullivan KJ, Griswold WG, Cai Y, Hallen B (2001) The structure and value of modularity in software design. In: ACM SIGSOFT software engineering notes, ACM, vol 26(5), pp 99–108
6. Shaw M, Garlan D (1996) Software architecture: perspectives on an emerging discipline. Prentice Hall, Englewood Cliffs
7. Shi S (2007) Use VB to process picture data of access database. China Sci Tech Inf 23:75

Integrating Simulation with Optimization in Emergency Department Management

Hainan Guo and Jiafu Tang

Abstract Nowadays, the hospital Emergency Department (ED) provides new challenges to their decision-makers because of high demands for services, high costs, and limited budget leading to inadequate healthcare resources. In this chapter, we integrated the simulation and optimization technique to address this important management problem in the ED, and we set up a decision support system to help the ED managers to design the optimal healthcare resources' staffing and scheduling plan in the specific environment. In an ED system model, many different categories of patients may require multiple services through a common sequence. Therefore, each service center has its own costs. The configuration of the resources will affect the overall efficiency of the system directly. So, our main work here is to set up a decision support system for the ED managers to design the optimal configuration of resources with the objective minimize patients LOS and minimize the overall costs separately.

1 Introduction

As we all know that, emergency department is a complex unit, which has many characteristics. For example, there are a lot of patients. It has a heavy rescue mission and so on. The patients who walk in the ED always want to get the treatment in time. However, most hospital ED budgets did not catch up with the demand for ED services made by growing populations and aging societies. So the reorganization plan needs to first address staffing and scheduling issues such as determining the correct size of the workforce. The quality of ED services has a close relationship with all the medicals' operations. Because of the patient overcrowding can not only

H. Guo (✉) · J. Tang
Systems Engineering, Northeastern University, Shenyang, Liaoning, China
e-mail: guohn403@sina.com

J. Tang
e-mail: jftang@mail.neu.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_126

affect the patients' psychological of treatment, but also can reduce medicals' quality of services. So, designing an optimal configuration plan within the limited resources has been one of the most important problems in the field of health care.

In order to solve the joint ED staffing and scheduling problem, in this chapter, we integrated the simulation and optimization techniques to address a management problem in the ED, and we set up a decision support system to help the ED managers to design the optimal healthcare resources' staffing and scheduling plan in the specific ED environment. Studies that did before have been carried out many research findings to help the ED managers to make decisions and to evaluate the efficacy and efficiency of their configurations [1, 2]. ED decision-makers must make the best decision within many constraints, such as limited budget, high demands for services, high costs and so on. In addition, they must submit the best configurations for alternative budgets in the different environments. So, attention is directed at providing insight into the system through what-if models, which can evaluate the alternative choices, as pointed out in [3]. The techniques used are most of simulation [4] and optimizations [5, 6], some studies combine these two techniques.

In the point of ED managers, the efficiency is measured by the cost of making a level of service, while in the ED patients' standpoint, except for receiving the high quality of service, and they also want to get the treatment in time. The most important reasons cause the patients having to wait are a lack of resources (i.e., not enough medical staffs or beds or the delay of the laboratory tests) and by their unbalanced availability. In fact, the budget can be allocated in the different ways to all kinds of ED resources, the unbalanced configurations will generate some resources are scare while others are abundant.

The scene considered here is an ED in the Xinhua Hospital in China, which had eight different resources with the different services required by four categories of patients. Patients of the same category go from one service to the next service in the predetermined paths, while different types of patients' paths are not quite similar. Each service is operated by one kind of resource, the type of resources here are the emergency nurse, registered nurse, surgeon, physician, pediatrician, bandaging staff, laboratory technician, and pharmacy nurse. The staffing and scheduling plans made by the ED managers for these resources were based on many conditions, and a good configuration could not only reduce the patients' LOS in the ED system, but also can maximize the use of given budget.

This chapter addresses the problem of how to select the optimal configuration of the ED resources. There are two objectives considered here in order to optimize efficiency, which are shown as following.

1. Determine the configuration of resources that minimizes the cost with the constraint on the patients' LOS.
2. Determine the configuration of resources that minimizes the patients' LOS with the constraint on the budget.

The procedures that we use interactively to solve above problems are: survey, system simulation, find the relation between inputs and performance, and establish optimization models.

As mentioned before, we integrated simulation and optimization techniques to solve this management problem. We used the simulation to reproduce the behavior of an ED system in order to analysis its performance and the outcome of different scenes. Optimization was used to obtain the optimal joint staffing and scheduling plans with above two objectives.

2 System Simulation

To accomplish the above works, a survey and a depth analysis of ED overall operations is the most critical prerequisite. In this chapter, we used three methods to collect the useful data, issuing questionnaires, discussion with ED managers, and communication with patients waiting in the ED system. Through talking with the ED managers and all categories of patients we can find that the most intractable problems are to minimize the cost in the ED managers' standpoint and reduce the LOS in the ED system in the patients' viewpoint. In addition finding out our objectives, these comprehensive surveys have been carried out in order to collect data on the patient arrival rates, the categories of patients, their distribution probabilities, the service times at each stage of the process, the transition probabilities and the transfer time between the different units. Table 1 gives the distributions of the service times at each stage of the process, and Table 2 presents the transfer time between the different units.

Through the survey, we find that ED is open 24 h 1 day and receives an average of 650 patients daily, including 5 % emergent patients, 43 % surgical patients, 41 % internal patients and 11 % pediatric patients. Patient arrival rates follow a process as shown in Fig. 1. The process begins when a patient arrives through the front door, and ends when this patient is released or sent to the ward.

Figure 2 illustrates the basic flow of the ED. The emergent patients were carried to the emergency room receiving the care directly, and then sent to the wards. On

Table 1 Resources' service time at the ED

	Distribution (min)	
Stage	First visit	Subsequent visit
Emergency nurse	T(10,15,20)	
Registered nurse	T(1,1.5,2)	
Surgeon	T(5.25,7.58,9.75)	
Physician	T(4.5,6.5,8.75)	T(0.5,2.5,4.75)
Pediatrician	T(5.08,7.75,9.83)	T(3.08,5.75,7.83)
Bandaging staff	T(5.5,7,10)	
Laboratory technician	T(4.75,5,5.5)	
Pharmacy nurse	T(1.5,2.08,2.98)	

Table 2 Transfer time between the different units

Transfer between the different units	Time (min)
Front door-emergency room	0.5
Register-surgical department	3
Register-internal department	5
Register-pediatric department	5
Surgical department-treatment room	1
Internal department-laboratory	1
Pediatric department-laboratory	1
Laboratory-waiting room	0.5
Waiting room-pharmacy	0.5
Pharmacy-internal department	3.5
Pharmacy-pediatric department	3

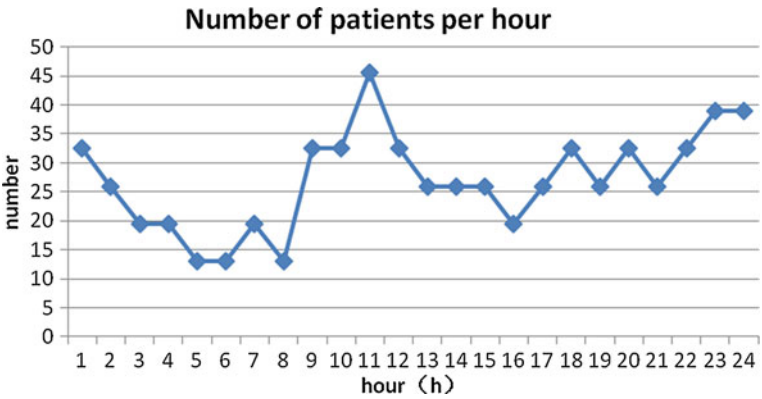
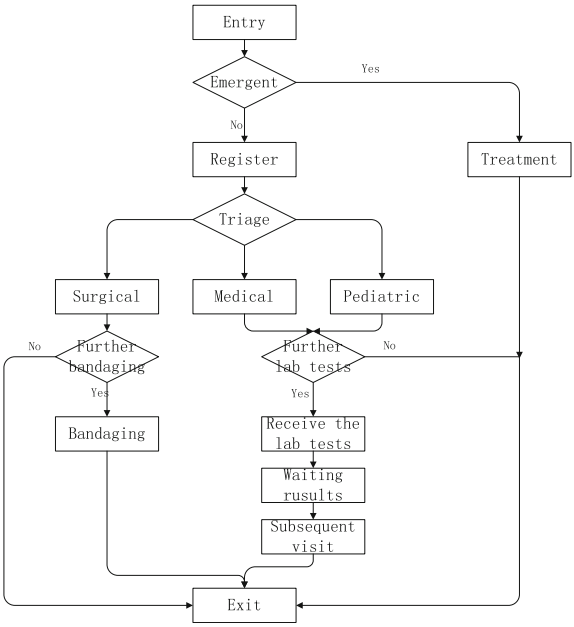


Fig. 1 Weekday patient arrival rate at the ED

the other hand, the walk-in patients first go to register, and then the surgical patient goes to the surgical department receiving the first visit. The surgeon will make a brief examination on a patient and decide whether if the patient would need further bandaging, which probably accounted for 70 %. Otherwise the patient is allowed to leave the hospital. It is the same with the surgical patients, the patients in the department of the internal medicine and pediatric, if the doctor decides the patients need further lab tests, accounted for 80 %, they go to the laboratory to receive the blood tests or any other lab tests. After that, the patients should go to the waiting room to wait their results, after a while, a nurse taking their test results will guide 40 % of them back to accept the subsequent visit. And the doctor decides whether a patient may leave or remain in the hospital.

It should be noted that, we must invite the ED managers to participate in the process of establishing the simulation model in order to make our model more feasible and valid.

Fig. 2 Basic flow of the ED



3 Find the Relation Between Inputs and Performance

The goal of this step is to find the relationship between the configuration parameters of the ED resources and the performance. In our study, the input parameters include patient arrivals λ and configurations x_{ij}, which means that the number of each type of resource in each service in each shift. And the performance considered by us here is the patients' LOS, y, gotten from simulation. We derive a set of observations $(\lambda^k, x_{ij}^k, y^k)$ from which the functional relation $y = f(\lambda, x_{ij})$ can be evaluated. Because of the nature of the underlying system, it is easy to find that the relationship between the inputs and the performance in our system is non-linear, so we must select the proper estimate method. In this chapter, we used radial basis function neural network (RBFNN) to solve this problem. Studies [7] that did have found that the RBFNN has simple structure, training speed and it can apply to multivariate non-linear regression model very well.

3.1 Training Set

In order to obtain a training set, the observations must generate only feasible configurations. For example, nobody on duty in any shift is an infeasible configuration. So, firstly, we should eliminate these infeasible configurations. And then, in

order to produce observations in the relevant regions of the parameter space, we have moved forward as follows: starting from the minimum number of patient arrivals per day, we defined 300, we kept simulating from the minimum number of servers in each service in each shift, and then increased the number of servers gradually until the optimal y predetermined was obtained. In the current patient arrivals, we randomly generated 10 groups of feasible configurations, and then simulated their patients' LOS. After finishing the above works, patient arrivals were increased and the whole processes were repeated until reaching the value of 700 for the daily patient arrivals. It should be noted that, the increment of daily patient arrivals in this chapter was 50 in each time.

3.2 Introduction of RBFNN

In this chapter, we resorted to using RBFNN [8], which was particularly flexible and effective in evaluating non-linear functions to find the relationship that linked the inputs and the performance. Due to RBFNN has been known to the broader nuclear community, so for brevity we omit some detailed descriptions of this technique.

RBFNN is a three-layer feed-forward network. It is defined by a number of input neurons, a single hidden layer and one or more output neurons. The structure of the RBFNN is presented in Fig. 3. The basic idea of the RBFNN is: it uses a RBF to be a "base" of the hidden layers, which can make the input vectors directly mapped to the hidden spaces without any weights. When the baricenters of the RBF are determined, this mapping will be determined. While the mapping relation from the hidden layers to the output spaces is linear, namely the output of the network is the sum of hidden layers' linear weights. The weights here are the adjustable parameters of the network. The estimating function is:

$$y_j = \sum_{i=1}^{h} \omega_{ij} exp(-\frac{1}{2\sigma^2} ||x_p - c_i||^2) \tag{1}$$

where $i = 1, 2, \ldots, h$ represents the number of hidden layer nodes, y_i is the jth practical output corresponding to the input samples, x_p is the input vector, c_i is the

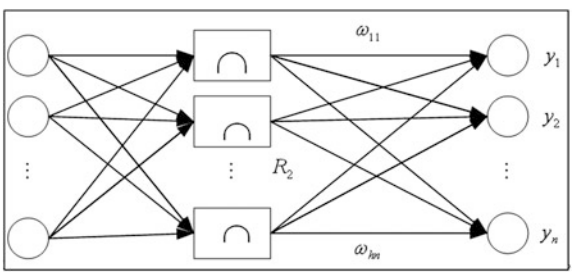


Fig. 3 Structure of RBFNN

Table 3 Analysis table of fitting results of RBFNN

Minimum error	Maximum error	Mean error	Mean absolute error	Standard deviation	Linear correlation
−0.491	0.537	0	0.083	0.123	0.999

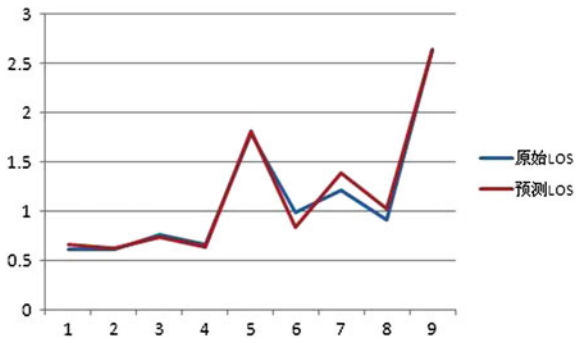


Fig. 4 Estimated and simulated values of LOS

*i*th baricenter, ω_{ij} is the weights from the hidden layers to the output spaces, and σ is the variance of the primary function.

The number of runs for each configuration was increased until the training of the network was sufficiently stable and presented generalization capability. With 20 runs for each configuration and a total of 270 observations, the training process was fully satisfactory and the results would not improve with more data. Table 3 presents the analysis of fitting results of RBFNN using the training sets were obtained used above method. And the Fig. 4 shows the fitting curve between the forecast results and the test samples. These answers state how the estimated function is able to capture the essential behaviour of the relationship, which we want to approach.

4 The Optimization Model

The main purpose of this chapter is to help the ED managers to design the optimal joint staffing and scheduling plans under the different constraints. In order to make a rational decision support system and give some useful suggestions for the ED managers, understanding and seizing of the problems are very important. So after making certain the relationship between the inputs and the performance, we can combine this function with some constraints to define the optimization models. The optimization problems considered in this chapter aims to two objectives, minimize the patients' LOS and minimize the total costs of all the ED resources, which have been mentioned in the introduction. In our system, all types of resources' working pattern are the traditional working strategy, i.e., 8 h strategy.

Table 4 The parameters design

Parameters	Design
$f(\lambda, x_{ij})$	Target function of patients' LOS
λ	Patient arrival rates
i	Type of ED resource
j	Shifts
n	Staffing of each type of resource
m	Total number of daily shifts
c_{ij}	The cost of each resource in each shift
C	Budget constraint
T	Patients' LOS constraint
$\lambda_{\min}$	The lower bond of patient arrival rates
$\lambda_{\max}$	The upper bond of patient arrival rates
L_{ij}	The lower bond of each type of resource
U_{ij}	The upper bond of each type of resource
x_{ij}	Number of each type of resource in each shift

Firstly, the collective parameters that be used in two problems are designed in the Table 4.

The specific optimization models are established as below:

4.1 Minimize LOS

In this sub-problem, our goal is to get the optimal configurations under the different patient arrivals with the objective that to minimize patients' LOS in the ED system. The constraints involve the budget restraints, the servers bonds and the different patient arrivals, which all obtained through survey. The specific model was established as follows:

$$min \sum_{i=1}^{n} \sum_{j=1}^{m} f(\lambda, x_{ij}) \tag{2}$$

$$s.t \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} x_{ij} \leq C \tag{3}$$

$$\lambda_{min} \leq \lambda \leq \lambda_{max} \tag{4}$$

$$L_{ij} \leq x_{ij} \leq U_{ij} \quad i = 1, 2, \ldots, n \quad j = 1, 2, \ldots, m \tag{5}$$

$$\lambda \in I, x_{ij} \in I \quad i = 1, 2, \ldots n \quad j = 1, 2, \ldots, m \tag{6}$$

Equation (2) defined the objective, to minimize the patients' LOS. Equation (3) was the first constraint should be satisfied, the resources' total daily costs must no more than the budget upper limit. Equation (4) presented the patient arrival rates must come within the upper and lower bonds. In a similar way, the Eq. (5) defined the number of each type of resource in each shift should in the scope of bonds given by the hospital. Through observing this model, we could find that all the constraints were the linear, while the objective was a non-linear function. Hereto, Matlab software had been used, and called the neural network toolbox meanwhile to solve this problem.

4.2 Minimize Cost

In this sub-problem, our goal is to get the optimal configurations under the different patient arrivals with the objective that to minimize the daily costs of all the ED resources. The constraints include the patients' LOS restraints, the servers bonds and the different patient arrivals, which all obtained through survey. The specific model was established as follows:

$$min \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} x_{ij} \tag{7}$$

$$s.t \quad \sum_{i=1}^{n} \sum_{j=1}^{m} f(\lambda, x_{ij}) \leq T \tag{8}$$

$$\lambda_{min} \leq \lambda \leq \lambda_{max} \tag{9}$$

$$L_{ij} \leq x_{ij} \leq U_{ij} \quad i = 1,\ 2, \ldots,\ n \quad j = 1,\ 2, \ldots,\ m \tag{10}$$

$$\lambda \in I, x_{ij} \in I \quad i = 1,\ 2, \ldots,\ n \quad j = 1,\ 2, \ldots,\ m \tag{11}$$

Equation (7) defined the objective, to minimize the resources' total daily costs. Equation (8) illustrated that the patients' LOS cannot exceed the given value, which determined according to the reality, especially the patient arrival rates. Apart from this, the others equations' meaning were same as the explanations introduced in the last part. It should note that, in this part, the objective was a linear function, while one of the constraints was non-linear, which were different from the last problem.

5 Computational Results

To solve these two sub-problems, we have test lots of account cases. Table 5 presents the results of optimization sub-problem 1 with the objective to minimize the patients' LOS. The results showed in Table 5 were all the optimal configurations of ED resources from the patients' LOS standpoint under the different budget constraints and patient arrivals. These results were all maximize the use of the budgets. In this chapter, we defined daily patient arrivals from 300 to 700, and the values of budgets ordered from 800 (RMB) to 2000 (RMB). For example, we could get the optimal resources' configuration that cost 1,398 and had the LOS of 1.13 h when the budget constraint was 1,400 and the patient arrival was 650. This must be the optimal configuration in the current constraints.

Table 6 shows the results of optimization sub-problem 2 with the objective to minimize the resources' costs. The results showed in Table 6 were all the optimal configurations of ED resources that used the least costs to hold the different patient arrivals within an acceptable patients' LOS. For example, when the upper bond of the LOS was 0.6 h and daily patient arrival was 450, we could get the optimal configuration with the least cost 1,079 (RMB) on the basis of no more than the upper bond of LOS constraint.

Through observing the Table 6, we found that in some cases, we could not get the feasible solutions when the daily patient arrivals were higher than capacity of the system under the too small LOS constraint, which was much easy to comprehend.

The results of above two tables are much meaningful. They can bring forth ED decision-makers many suggestions from the perspective of the patients and the managers of the hospital separately. We established a decision support system for them, which could not only help the ED managers to design the optimal configurations within the different constraints, but also could guide them to optimize the overall operations of ED by adjusting and making advisable policies. The main signification of this work was to propose four advices to the ED managers that help them to optimize the ED system, which was specifically introduced as follows:

1. This support system can help the ED managers to choose the most reasonable configurations that they consider, based on the realities and the balanced state they hope to achieve.

For example, through observing Table 5 we can find that patients' LOS in the ED system is 1.4 h when the daily cost constraint is 800 and daily patient arrival is 450, while the patients' LOS is even as high as 5.08 h when the daily patient arrival rises to 500. However we find that when the cost constraint rises to 1,100, patients' LOS will reduce to 1.29 h in the case of 500 of daily patient arrival. This phenomenon illustrates that the cost constraint of 800 can only satisfy the daily patient arrival does not exceed the 450, and the hospital needs to increase the budget properly when there are more patient arrives. Different realities will get different results, and one of the contributions in our decision support system is to help the

Table 5 Optimization results with the objective of the minimize patients' LOS

Patient arrivals (per day)	Budget constraints in RMB									
	800		1,100		1,400		1,700		2,000	
	Cost	LOS (h)	Cost	LOS (h)	Cost	LOS (h)	Cost	LOS (h)	Cost	LOS (h)
300	800	0.65	1,098	0.58	1,338	0.56	1,338	0.56	1,338	0.56
350	800	0.70	1,098	0.60	1,338	0.57	1,338	0.57	1,338	0.57
400	799	1.01	1,100	0.89	1,400	0.79	1,612	0.56	1,612	0.56
450	799	1.40	1,100	1.22	1,400	0.99	1,612	0.57	1,612	0.57
500	798	5.08	1,096	1.29	1,400	0.82	1,697	0.56	1,765	0.56
550	798	6.53	1,096	1.60	1,400	0.95	1,697	0.57	1,765	0.56
600	798	7.30	1,096	2.25	1,400	1.07	1,697	0.57	1,765	0.56
650	784	7.70	1,094	2.87	1,398	1.13	1,700	0.58	1,995	0.56
700	784	9.66	1,094	4.04	1,398	1.41	1,700	0.60	1,995	0.56

Table 6 Optimization results with the objective of the minimize resources' cost

Patient arrivals (per day)	Upper bonds of patients' LOS (h)									
	0.5		0.6		0.8		1.0		1.2	
	Cost	LOS	Cost	LOS	Cost	LOS	Cost	LOS	Cost	LOS
300	1,290	0.50	919	0.60	654	0.69	654	0.69	654	0.69
350	1,320	0.50	943	0.60	733	0.80	654	0.83	654	0.83
400	Infeasible		1,079	0.58	886	0.80	780	1.00	694	1.12
450	Infeasible		1,079	0.59	911	0.80	910	1.00	807	1.13
500	Infeasible		1,182	0.59	1,098	0.79	985	0.99	934	1.17
550	Infeasible		1,305	0.59	1,203	0.80	1,098	1.00	950	1.19
600	Infeasible		1,406	0.59	1,298	0.80	1,153	1.00	1,058	1.18
650	Infeasible		1,521	0.59	1,413	0.80	1,313	0.99	1,235	1.17
700	Infeasible		1,680	0.60	1,579	0.80	1,450	1.00	1,380	1.20

ED managers to deploy the budget standard and the reasonable configurations properly in different daily patient arrivals.

2. This support system can help the ED managers to weigh the acceptable of actual budget and the patients' LOS standard from the perspective of the hospital and patients standpoint separately.

For example, we can find that the total wage costs of ED resources will be an obvious rise when the patients' LOS standard is reduced a little by observing the Table 6. However, there is not a strait requirement of the upper bond of patients' LOS in reality. So, the ED managers can appropriate relax the standard of patients' LOS in the ED system, which can greatly reduce the total costs of ED resources.

3. This support system can help the ED managers to re-optimize the medicals' configurations in order to make the system better.

For example, we all know that all the EDs have their own original configurations, and we can calculate their wage costs, we assume to be 1,350 (RMB). And based on the historical, the ED managers can estimate the daily patient arrival is 500. After that, we can get the patients' LOS in the ED system by using the RBFNN. We setting C = 1,400, $\lambda_{min} = \lambda_{max} = 500$, and we will get a new configuration at the least patients' LOS through running optimization Problem 1. This new configuration's total cost will not exceed the existing cost, and it is the optimal choice, which can make the patients' LOS lowest.

4. This support system can help the ED managers to maximal use the existing conditions to design the optimal configuration of ED resources.

It is very important to note that only to revise the budget and determine the patients' LOS standard based on the current situation is not enough. ED managers always need to design the proper configurations of all the resources. But sometimes, how to deploy the configuration is most troublesome problem. In this chapter, this problem has been solved. ED managers only need to give the constraints, and then through running these two optimization models they will obtain the optimal configurations directly.

6 Conclusions

This chapter designed a decision support system for the operation of the ED by integrating simulation with optimization. We presented a methodology, which used system simulation combined with optimization to determine the optimal joint staffing and scheduling plans, called configurations, to minimize total costs from the standpoint of hospital and to minimize the patients' LOS in the ED system in the perspective of patients. In addition, a decision support system was designed to help ED managers to either evaluate different situations of configurations.

References

1. Broos M, Mario V (2010) Branching strategies in a branch-and-price approach for a multiple objective nurse scheduling problem. J Sched 13:77–93
2. Broos M, Mario V (2012) An integrated nurse staffing and scheduling analysis for longer-term nursing staff allocation problems. Omega 12(4):1–34
3. Bakircioglu H, Tocak T (2000) Survey of random neural network applications. Eur J Oper Res 126(2):319–330
4. David S, Yariv M (2005) Emergency department operations: the basis for developing a simulation tool. IIE Trans 37:233–245
5. Francis DV, Otis B (2011) Jennings. Nurse staffing in medical units: a queuing perspective. Oper Res 6(59):1320–1331
6. Haykin S (1994) Neural networks. Macmillan Publishing Company, Englewood Cliffs
7. Jinn-Yi Y, Wen-Shan L (2007) Using simulation technique and genetic algorithm to improve the quality care of a hospital emergency department. Expert Syst Appl 32:1073–1083
8. Navid I, Dave W (2012) Setting staffing requirements for time dependent queuing networks: the case of accident and emergency departments. Eur J Oper Res 219:531–540

RUL Assessment and Construction of Maintenance Strategies for Engineering Objects

Alexander Khodos, Aleksandr Kirillov and Sergey Kirillov

Abstract Actual trends in the construction of systems of various types of maintenance; rather their choice in a particular situation is usually determined by the technical state of engineering at the time. Accurate information about its operating condition and prognosis of all detected trends to the side of failures or increasing risks or failures are necessary for the correct determining the optimal maintenance strategy and in general strategy of operational service of engineering object. Thus the life cycle of engineering relatively the types of operational service and maintenance methods is divided into several time intervals determined by the conditions of the object and prognosis. Their maintenance sets of measures are determined at each time interval from operating measures of self-maintenance to end of life management: reuse-remanufacture-recycle. Choice of various types of maintenance is more effective at accurate determining the state of engineering at the failure progression timeline. Account of optimization problem of operating costs throughout the life cycle of engineering in many respects is determined by exact definition of engineering positions on the failure progression timeline and the rate of evolution to failure state of engineering at particular time interval. Thus, different types of monitoring for engineering are required for the construction of systems of local RUL assessments. The considered problems of the state and RUL assessments are solved by integration of remote computing PHM clusters and on-board diagnostic systems of engineering object. RUL evaluation methods in the monitoring process at all time stages of the engineering operating are demonstrated in this chapter. Maintenance strategy for all operating phases and some methods of self-maintenance available for use by varying the control parameters of an engineering object are discussed. In particular, the management solutions in order to minimize maintenance costs for certain time intervals of engineering object operating are given.

A. Khodos (✉) · A. Kirillov · S. Kirillov
SmartSys Prognosis Center, Godovikov Str, 9, Building 25, Moscow, Russia
e-mail: smarttechappl@gmail.com

A. Kirillov
e-mail: smarttechappl@gmail.com

S. Kirillov
e-mail: skirillovru@gmail.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_127

1 Introduction

The basis for determining the optimal maintenance strategy is the information about the technical state of engineering at the current moment and accurate assessment of the evolution of technical condition in the future (prognosis). In order to determine the strategy maintenance information on the impact of management parameters of engineering on the change of technical state at the moment and change of the prognosis parameters provided changing management parameters is needed. Based on received information hereinafter temporal assessment of the development of early or hidden signs of degradation and failure are necessary. Based on the temporal assessments it is possible to optimize various types of maintenance: condition Base maintenance, predictive maintenance, and self-maintenance. In previous chapters [1, 2] the authors have described the models, establishing temporal hierarchy of all possible sets of predictors of degradation and failure. In accordance with these PHM hierarchical models, the system state is defined by a vector consisting of the sequence of observed values of sensor readings for a fixed time interval. Sequential change of values of the state vectors in the operation of engineering determines trajectories in a multidimensional vector space.

From the hierarchical model [1, 2] follows:

- the set all permissible trajectories is divided into classes. Each class defines a time interval. Union of all time intervals is the life cycle of the mechanism;
- each class of trajectories is characterized by its features (predictors). Predictors of one class in reverse time are degenerate relative to predictors of the next class. Themselves predictors are entropic, topological, stochastic characteristics of the trajectories, as well rational expressions of moments of multidimensional distribution functions;
- within its class predictors are varies, characterizing the degree of proximity to the boundary of class trajectory. The changes of predictors temporally are determined by evolution equations of the process for the density of the transition probabilities;
- estimate of RUL(K), K = 0, I, II, III, IV for each trajectory in its class are determined from evolution equations represent time to achieve the boundaries of class. The sum of all classes of RUL (K) is equal to the life cycle of mechanism;

II Incipient of fault hidden cause;
II Suddenly of fault cause и Incipient of fault cause;
III Suddenly incipient fault и early incipient fault;
IV Component or subsystem of failure.

Sections 2 and 3 of this chapter focuses on models and algorithms for determining the RUL (III) in class III. Model of RUL (III) calculations is constructed on the representation of the probability function of transition from one vector state to another. The transition probabilities are represented by Feynman path integral. The

evolution equations for the transition functions are determined by the Feynman representation and moments of the density of transition probabilities are calculated. In some cases, the moments are determined analytically, that greatly simplifies the experimental verification of the results and provides a basis for the formalization of the optimization problem of maintenance. The model can be extended to classes I, II by the reformulation of Feynman path integral on manifolds and homogeneous spaces of the groups of the state symmetry.

Also there are discusses the model of change of the trajectories of subclass "Suddenly incipient fault" of class III. Connection between changes of characteristic features of the trajectories induced by noise, transitions and critical phenomena in condensed matter physics is traced.

The constructed analytical RUL estimates allow to determine the optimal strategy of maintenance and basic principles of monitoring. Well as the results enables to determine the basic parameters for the remote monitoring and rules of construction of the chronological databases.

Section 4 discusses the mathematical formulation of optimization problems of maintenance strategy and ROI estimates.

The material is largely descriptive. Since the purpose of the article is to cover the optimization problem of maintenance as a whole and to determine in its special role of remote PHM monitoring for the calculation of RUL in the context of optimization of maintenance strategies.

2 Models of RUL Calculating

Model of PDF representation for the transition probabilities of vector processes defined on the set of the wavelet coefficients of the observed vibration signal in the form of Feynman path integral is taken as a basis for estimates of RUL (III) of internal combustion engines, reciprocating mechanisms, transmissions, clutches and bearings [2]. After the procedure of secondary discretization, signal s represent as a vibration signal in an angular variable $\boldsymbol{\varphi}$, $S'(\boldsymbol{\varphi})$, $\varphi \in [0, 4\pi]$.

At the next step a signal is represented by a set of finite segments $\left\{{}_{0}^{N}w_{j,k}\right\}$ of wavelet coefficients $S'(\boldsymbol{\varphi})$ as follows. A set of finite segments of wavelet decomposition of vibration signals is

$$\{R_i\}\underline{\underline{\text{def}}}\left\{{}_{i}^{N}w_{j,k} : iN^* \leq N \leq (i+1)N^*, i = 0, 1, 2, 3\ldots\right\}$$

at fixed scale j and translation k, where (j, k)—the index of wavelet decomposition coefficients and N-number of cycle of engine crankshaft.

The model there are used models constructed on representation $P(\boldsymbol{R}, i)$ probability of transition for the i steps to the state $\boldsymbol{R}$ in the form of Feynman path integral

$$P(\boldsymbol{R}_0, \boldsymbol{R}_L, L) = \int_{\boldsymbol{r}(0)=\boldsymbol{R}_0}^{\boldsymbol{r}(L)=\boldsymbol{R}_L} \mathcal{D}[\boldsymbol{r}(s)] \exp\left[-\int_0^L ds < \xi_L > \boldsymbol{r}_L^2(s) \right]$$

L- is continuous analog i, in $P(\boldsymbol{R}, i)$; $\langle \xi \rangle$—is the average value, $\Delta \boldsymbol{R}_i$; $\boldsymbol{r}_L^0(s)$—is parameterization of polygonal $\{\Delta \boldsymbol{R}_i; i = 1, 2, 3\ldots\}$.

Representation of the transition probability from one physical state to another by means of Feynman path integral in various physical models developed in [3, 4]. Let's use the same approach modified under the high dimensions N^* space of finite segments of the wavelet coefficients. Such representation of the transition probability leads to partial differential equation for $P(\boldsymbol{R}_0, \boldsymbol{R}_L, L)$. The obtained equations allow calculating the transition probability in an explicit form or calculating moments of the transition probability. The solution of inverse equations for the moments gives values L or discrete values i, that irreversible for transition to the pre-fixed state $\mathbf{R}$, which is the basis for RUL estimate.

The set of wavelet coefficients $\left\{ {}_0^N w_{j,k}; N = 1, 2, 3\ldots \right\}$ of signal $S(\varphi)$ is transformed into a set of finite segments of the wavelet coefficients multidimensional vectors of state of the system $\boldsymbol{R} \in R^{N^*}$ with a certain fixed dimension N^*.

The value N^* is definite by formula

$$N^* = \min_N \left(\left| PDF\left(N - \check{S}\right) - PDF(N) \right| < \varepsilon \right)$$

with a predetermined value ε, defined by taking into account the accuracy of the measuring sensors and digital electronics, $\check{S}$—the minimum integer starting from which the previous estimate is valid.

On a set of sequential finite segments is determined by process

$$\Delta R_i = R_i - R_{i-1}.$$

As a result the process of random walk vector $\Delta \boldsymbol{R}_i$ in N^*—dimensional space is considered and the problem of prognosis and assessment of RUL (III) is reduces to calculating the number of steps i, for which the state vector $\boldsymbol{R}$ gets from the start state $\boldsymbol{R}_0$ in the class III with probability $P(\boldsymbol{R}, i)$, transition from early incipient fault to component or subsystem failure, if $\boldsymbol{R}$ is state of boundary of class III. RUL estimates largely depend on the stochastic properties of the observed signal. Taking into account this dependence from the Feynman representation of the transition probabilities the following analytical expressions for RUL follow.

1. Models of the free random walk.

$$RUL(III) = \frac{< \left\| \boldsymbol{R}^{\boldsymbol{C}} \right\|^2 >^2}{(N^* - 1) < \xi >} \tag{1}$$

L—is the number of revolutions required to achieve the state vector $\boldsymbol{R}$ with root-mean-square norm $\left\langle \|\boldsymbol{R}\|^2 \right\rangle^0$

$\boldsymbol{R}^{\boldsymbol{C}}$—belongs to the boundary of transition III–IV (IV—class component or subsystem failure).

2. Models of random walk with constraints. Under condition of confirmation of hypothesis of model of random walk with constraints, the estimation of RUL (III) is determined as solution of equation

$$\langle \|\boldsymbol{R}\|^c \rangle^2 = 2\left\{ \Theta \mathrm{RUL(B)} - \Theta^2 \left[1 - e^{-\mathrm{RUL(B)}/\Theta} \right] \right\},$$

$$\Theta = \frac{const}{\langle \xi \rangle (N^* - 1)}$$

3. Models of random walk in a non-simply-connected domain. Under condition of confirmation of a hypothesis of model of random walk in a non-simply-connected domain estimates of RUL(III) is determined as solution of equation

$$\langle \|\boldsymbol{R}\|^c \rangle^2 = \langle \xi \rangle^{1/(N^*+2)} \left(\frac{N^* + 2}{3} \sqrt{\frac{2\langle \xi \rangle}{N^*}} \mathrm{RUL(B)} \right)^{6/N^*+2}$$

3 Catastrophes and Non-Gaussian Distribution

Having determined the nature of the stochastic process, the evolution equation of the probability of transition from one state to another can be determined within the class of trajectories or states. Based on the evolution equation the prognosis of the development is possible provided the immutability of the properties of the stochastic process. The problem of monitoring system is reduced to the control of the stochastic properties of the process. However, the proposed models do not always take into account the entire existing situation in practice. Accounting of this fact entails a more subtle classification of classes of hierarchical model.

Therefore, the further separation of class III is illustrated by the following example. Let's return to the case of free walking of the state vector. Evolution equation in this case is the classical diffusion equation and the formula for estimating the RUL has the form [1]. In the derivation of the evolution equation it was assumed that the density of the distribution function ξ has a Gaussian form. But what happens to the process when the density of the Gaussian distribution ξ is violated? Among other things, even against the background of stationary solutions of evolution, the sudden transformation of the distribution function described by the theory of catastrophes can happen.

Different scenarios for the probability density of the basic processes are possible. At stationarity of the base process, the probability density is the solution of the stationary evolution equation. However, changes in the external and internal parameters of the mechanism can lead to the transformation of the stationary solution and consequently affect the assessment of RUL. Such changes can have different physical nature, but for the purpose of PHM monitoring is sufficient be able to calculate all the parameters from the observed signals.

Changes in RUL assessment are possible in cases of transformation of the stationary solutions of evolution equation. This process is called "noise-induced transition" and the basis for the modeling of such transitions are set out in the work [5]. In the simplest case, the probability of transition is a solution of the stationary Fokker–Planck equation, and the external noise is additive with the intensity σ, the transition probability is represented as

$$P(\boldsymbol{R}) = const \cdot \exp(U(\boldsymbol{R}, \lambda, \sigma))$$

λ—the internal parameter, σ—the intensity of the external noise.

This representation allows to describe the sudden restructuring of transition function $P(\boldsymbol{R})$at slow variation of the internal parameters λ, σ. For the parameter problem, such transitions are described by the simplest bifurcation sets or catastrophes. Changes in the parameters of external noise and internal parameter shown as continuous paths on Fig. 1. As the figure shows in the neighborhood of bifurcation sets of reconstruction of the process happens, that is, the distribution function sharply changes its character, for example, becomes bimodal. In other cases, this model defines the so-called loss of stability of the zero sign. The transition is described, when the initial distribution density had a δ-shaped carrier in the neighborhood of zero. Then, when changing intensity of the external noise and

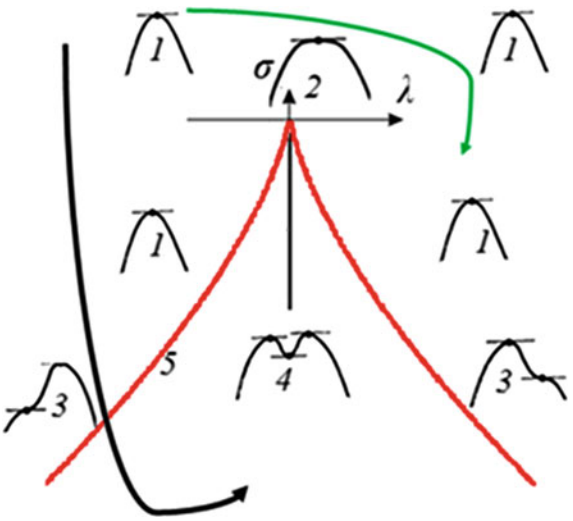


Fig. 1 Model of transformation of probability distribution function; *red line* bifurcation set, *green* and *black lines* trajectories of parameter change, *1–5* distribution function

management parameter there was a transformation $P(\boldsymbol{R})$at which the zero sign has lost its stability and the distribution function $P(\boldsymbol{R})$ has acquired non-zero moments.

From here, an important conclusion about the necessity of monitoring of all physically justified wavelet coefficients follows, if it is talk about vibration.

In practice, the loss of stability of the zero sign means the generation of additional high-frequency noise that always indicates the appearance of a new process in friction pairs, in the process of combustion of the fuel mixture, in the gearbox, transmissions, etc. To optimize the process of monitoring is sufficient to monitor the parameters of external noise and to management parameters to order to determine their proximity to the bifurcation sets. If thus possibility of changing the value of parameter λ remain, then thereby, there is an additional opportunity for self-maintenance, which reduces to monitoring and changes of the parameter λ in dangerous changes of the intensity of external noise σ. Complications of model are related to the increasing number of management parameters and nature of external noise. In the case of more than two parameters of the problem, the bifurcation set becomes complex hyper-surface with self-intersections in a multidimensional space. For example, if three parameters, one of the sets bifurcation is shown on Fig. 2 (dove tail). A more detailed classification of bifurcation sets can be found in work [6]. If the external noise is multiplicative, the above scenario is complicated, because minimums of the potential do not coincide with the maximums of the most probable values of the transition function. For the application of the theory of catastrophes, in this case it is necessary to use more sophistic models.

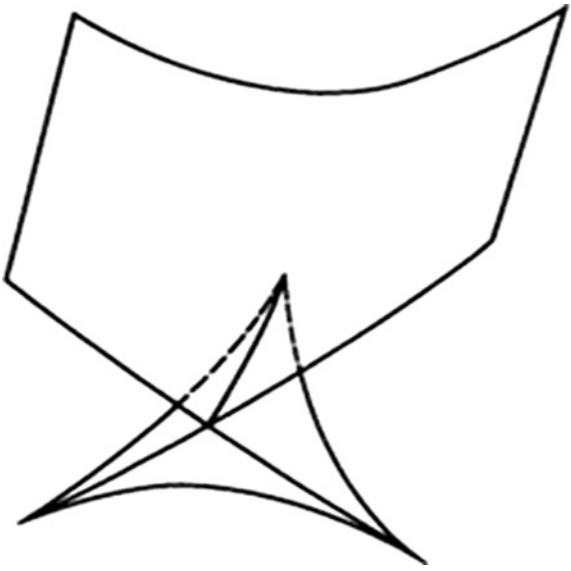


Fig. 2 The catastrophe "dovetail"

But some interpretations are possible in the following sense. Multiplicativity of external noise "smears" bifurcation picture of transformation $P(\boldsymbol{R})$, therefore described transitions are interpreted as mechanisms of transformation of one types of probability densities of the transition to the other.

The loss of stability of the zero sign leads to the generation of a new process and its further development, evolution. Formulas for RUL estimates vary in mind the changed nature of the process. In particular, the representation of the propagator in the form of Feynman path integral, derivation of evolution equations are used for estimates.

Interest are the rapid transitions or slow evolution to heavy-tailed distributions. In this case, when the transition on two possible mechanisms has occurred, the further evolution of the transition probability density will be described by the equations in fractional derivatives. The evolution equation of the diffusion type transforms into the equation with fractional derivatives having the following form [7]:

$$\frac{\partial^{\beta}\boldsymbol{P}}{\partial t^{\beta}} = \frac{\partial^{\alpha}}{\partial|x|^{\alpha}}(\mathcal{A}\boldsymbol{P})$$

$0<\beta\leq 1, 0<\alpha\leq 2$—critical exponents, $\mathcal{A}$—analog of the diffusion coefficient.

The evolution equation is called Fractional Fokker–Planck–Kolmogorov Equation (FFPK) or simply fractional kinetic equation (FKE). For $\beta = 1$ and $1<\alpha<2$ the FKE corresponds to the Levy process.

In particular, Levy processes have a radial part in the form

$$P(\|R\|) \sim (\|R\|)^{-(1+\alpha)}$$

If α decreases, then $P(\|R\|)$ increases at $1<\|R\|<\infty$. That is, at α, close to zero in the trajectories of the process is dominated by large jumps, if α is closed to 2, then the process is moved small jumps. The mathematical expectation at $\alpha<1$ is infinitely. Thus, the transition of the system to the processes, which have the distribution function with heavy-tailed distribution, generates unfavorable regimes as discussed above Levy flights. Discontinuous changes as noted in the work [8] cause strong mechanical overload of engineering and as consequence to the generation and development of micro-cracks, wear of the antifriction layer of friction pairs, changing the hydrodynamic lubrication regimes in plain bearings, i.e. the degradation of the material.

Moreover, the total damage estimated for example as the total length of microcracks at the time instant $N\Delta T$ less damage acquired in the next time interval $N\Delta T + \Delta T$. Non-linear growth of the total damage is the result of the exponential decrease of the radial distribution function [9].

Thus, many of the above methods are based on an analysis of random walk of the state vector on a multidimensional lattice or in the continuum limit of the random walk in a multidimensional space. The simplest form of random walk is the walk with a Gaussian distribution function. In this case, the process has diffusive

nature, which simplifies the RUL estimations, because the class of diffusion equations is well studied. However, in the process of normal wear of materials the processes of random walk are becoming more complex. In these management problems of trajectory of process receives special complexity. The increase of the complexity of random walk nature are not exempt from the need to RUL estimates in determining the maintenance strategy. Developed random walk model is the basis for estimates of RUL in all known cases of possible scenarios of the process. For example, in the prognosis problem it is actual the case of the trajectory return to the start state vector and assessment of the number of engine cycles required for such return.

Here are the following formulas [4]:

$$P(l,t) = \sum_{n=0}^{\infty} p_n(l) t^n$$

$$p_n(l) \sim \frac{1}{(2\pi n)^{\frac{d}{2}}} e^{-\frac{l^2}{2n}}$$

$$P(0,1) = \sum_{n=0}^{\infty} p_n(0)$$

$P(l,t)$—generating function for the transition probabilities,

$p_n(l)$—the likelihood of achieving a node l for n steps, discreet analog $P(R_0, R_L, L)$

Finally, it obtain

$$P(0,1) = \sum_{n} \frac{1}{(2\pi n)^{\frac{d}{2}}}, \quad p_n(0) \sim n^{-\frac{d}{2}}$$

Evaluation of the probabilities of return to the start state is very important for the following reason: it provides valuable prognostic information determining the degree of removing the state of the mechanical system from the etalon or start by elongation of the return trajectories. This, in turn, allows to estimate the probability of crossing the boundaries of class, after which the state of the system becomes irreversible in the context of return to the initial state. On the other hand, estimate of the length of cyclic trajectories enables to give the most accurate estimate of the Kolmogorov complexity of the trajectory and the modified Shannon entropy that it is important for the early prognosis [10, 11]. As well, assessment of the topological entropy of the dynamical system follows from the characteristics of cyclic trajectories [12]. In turn, the entropy characteristics and their evolution are early predictors of failures [1].

On the other hand, the random walk model used here for the prognosis is closely related to the theory of critical phenomena, and therefore the renormalization group method, the method of ladder diagrams, methods of quantum field theory [4, 13].

Such relationship allows to use apparatus of the above modern physics models for the effective evaluation of the predictors and their evolution.

These analogies are useful for calculating the critical exponents β, α for determining degree of proximity of mechanical system to areas of failures using thus apparatus of the theory of critical phenomena of condensed matter. Then there is the opportunity to explore the evolution of the finite segments in small as well in large dimensions. In particular, to estimate the correlation lengths in the neighborhood of noise-induced transitions are useful formula

$$\xi_c \cong const \cdot |\varepsilon|^{\nu} (\varepsilon \to 0)$$

here ε determines a neighborhood of the critical values of the external noise, ν—the critical exponent.

The increasing complexity of the random walk processes is also associated with random walk without self-intersections. This situation is possible when the self-intersection of trajectories impede physically irreversible processes in exploited engineering. Account of forbidden states that cuts forbidden regions in the multi-dimensional space of states also leads to the tasks of the random walk with limitations.

The problem of self-maintenance is essentially the problem of determining the time dependence of the management parameters in order to avoid the trajectories which lead the system to forbidden states, and that means that in the problem of the random walk the limitation appears in the form of some interaction of state with the forbidden regions representing repulsion. However, this interaction changes the RUL estimates. Account of physically irreversible processes of engineering also leads to the necessity to introduce a pairwise interaction between states. Detailed character of such interactions is determined through the identification of possibility of statistical interaction between states and requires continuous monitoring.

Returning to the noise-induced transitions, it is need to add the following. RUL estimates in the case of additive noise are replenished models from the theory of critical phenomena in condensed matter physics. Given utterance confirms the following paradigm of development of remote monitoring systems: namely, the need for a more precise determination of the state of engineering, a constant flow of telemetry to the remote server from different classes and types of engineering objects. Such flow is necessary on the one hand for determining basic characteristics of stochastic processes of observed sensor signals, and construction more and more early predictors of degradation and failures. On the other hand, the existing for today methods is sufficient for recoupment of the pilot prognosis systems. That is, the prognosis system can be developed only on condition that functioning on the market of maintenance services, obtaining the necessary statistical data from operating engineering for the development of models of the entire earlier prognosis.

4 Strategy of Maintenance and ROI

Thus, the set of possible states of engineering is divided into classes. The set of classes has hierarchical structure. Evolution of the states during operation is determined by the trajectory of the states. Predicting the evolution of the trajectory is reduced to the determination the properties of the stochastic process or the trajectory, then to the determination of evolution equations from which time-dependent characteristics of the trajectory, entropic, dimensional, cycling, etc. are determined. Dedicated characteristics are predictors of signs of failure, the cause of damage and degradation. For early prognosis, the predictors determine types of trajectories being on which the system achieves most rapidly regions of failures. The problem of self-maintenance is reduced by using the management parameters of the system to the choice of optimal trajectories, that is, those trajectories on which the system is within its class the maximum time. In all cases, the locally determined value of RUL is the subject of optimization. Determination and optimization of RUL is necessary also to optimize maintenance costs and determine the maintenance strategy. Just RUL is needed to determine return on investment.

5 Conclusion

Thus, the scheme of hierarchical levels is complicated. Accounting of rapid change adds to the previous scheme the additional subclasses of parameterized trajectories or states, on which the catastrophic changes in the distribution functions of processes, are possible. Such processes as Levy flights, depending on the values contained in them parameters, lead to the observed jumps, in turn, causing irreversible processes of accelerated wear, avalanche formation of microcracks, etc. To estimate the time of such processes can be under condition of continuous monitoring, constantly calculating the α, β parameters, i.e. estimating the evolution of the distribution function and comparing it with solutions of evolution equations with fractional derivatives. However, the degradation of the material will also take place at stationary processes with a heavy-tail distribution. If such processes are still registered, then task is the assessment of the degree of maintainability

Assessment of ROI is useful base on assessments of RUL in each hierarchical class. This allows to compare the cost of maintenance measures, the total cost of the monitoring system. At occurrence of states with heavy tails distributions, the damage of engineering is the maximum values in vector processes. This means that caused damage in case of its non-linear growth is apparently higher than the total damage to other stages of the operation. Perhaps this observation will formalize the calculation of ROI and develop them like methods in actuarial calculations. It becomes obvious to attract more and more sophisticated mathematical and physical theory to the RUL estimates. It is clear that over time, this trend will continue to increase in view of the need the earliest and more accurate prognosis.

References

1. Kirillov A, Kirillov S, Pecht M (2012) The calculating PHM cluster: CH&P mathematical models and algorithms of early prognosis of failure. In: Conference on prognostics and system health management (PHM 2012, Beijing, China), doi:10.1109/PHM.2012.6228771
2. Kirillov S, Kirillov A, Kirillova O (2011) Theoretical models and market architecture of PHM monitoring systems. In: Prognostics and system health management conference (PHM-2011 Shenzhen) May 24–25, China
3. Kleinert H (2004) Path integrals in quantum mechanics, statistics, polymer physics, and financial markets. World Scientific, Singapore
4. Ziman JM (1979) Models of disorder. Cambridge University Press, Cambridge
5. Horstemke W, Lefever R (1984) Noise-induced transition, theory and applications in physics, chemistry, and biology. Springer, Berlin
6. Arnold V, Varchenko A, Husein-Zade S (1982) Singularities of differentiable mappings. Classification of critical points, caustics and wave fronts. Nauka, Moscow (in Russian)
7. Zaslavsky G (2002) Chaos, fractional kinetics, and anomalous transport. Physics reports
8. Kirillov S, Klubovich V, Vagapov I (1985) Focusing of elastic pulse in solids with dislocation cluster. Izvestiya AN BSSR (in Russian), p 6
9. Malinetskij G, Potapov A (2002) Modern problems of nonlinear dynamics. URSS, Moscow (in Russian)
10. Kirillov S, Kirillov A, Kirillova O (2011) System of the automatic preventive on-line monitoring and diagnostics of car engines on the basis of the new methods of preventive diagnostics. SAE world congress and exhibition, Detroit, USA, 2011 Technical paper 2011-01-0747. doi:10.4271/2011-01-0747
11. Ray A (2004) Symbolic dynamic analysis of complex systems for anomaly detection. Sig Process 84:1115–1130
12. Martin N, England J (1981) Mathematical theory of entropy, England encyclopedia of mathematics and its applications, vol 12. Addison-Westley Publishing Company, Reading
13. De Gennes P (1979) Scaling concepts in polymer physics. Cornell University Press, Ithaca

The Problem of PHM Cloud Cluster in the Context of Development of Self-maintenance and Self-recovery Engineering Systems

Aleksandr Kirillov, Sergey Kirillov and Michael Pecht

Abstract The target problems of PHM computing cluster for different types of maintenance: condition-based maintenance (CBM); predictive maintenance (PdM); self-maintenance and self-recovery, are discussed in this paper. All types of maintenance are determined by the current state of engineering, degree of wear, the duration and conditions of operation. Therefore, various types of maintenance define various financial and time costs. Recent trends in the development of different maintenance strategies are aimed at the creation of self-maintenance and self-recovery engineering systems. However, to support such systems are required new models and methods of determining the technical state of engineering and its prognosis. That is, each of the stages of engineering object maintenance must be supported by appropriate methods of diagnosis of the condition of engineering objects, methods of accurate prognosis and assessment of time intervals of prognosis reliability. Consequently, the problem of supporting various types of maintenance and the development of appropriate formalisms, methods and algorithms to analyse the condition of object and prognosis for each type of maintenance should be included in the PHM problems. PHM problems should represent the universal concept and system of algorithms and rules, capable also to an estimation of efficiency of chosen strategy maintenance, minimization of cost and reduce operating costs. These necessary methods should diagnose condition at all operation phases and estimate time of achievement of borders of each operation phase. Thus, PHM problems and determining the time hierarchy predictors of engineering conditions become the base for the development of maintenance and self-maintenance, but not only that. The paper is a review of actual problems for PHM in the context of the practice of application of computing clusters, algorithms and scenarios for the

A. Kirillov (✉) · S. Kirillov (✉)
SmartSys Prognosis Center, Godovikov Str, 9, Building 25, Moscow, Russia
e-mail: smarttechappl@gmail.com

S. Kirillov
e-mail: skirillovru@gmail.com

M. Pecht
Center for Advanced Life Cycle Engineering, University of Maryland, College Park, USA
e-mail: pecht@calce.umd.edu

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_128

organization of the global system of development of self-maintenance systems based on of computing clusters are specified. This review is based on testing the PHM hierarchical models and algorithms for the pilot version of the PHM cluster at the analysis of failures of internal combustion engines and mechanisms of high complexity.

1 Introduction

Actual trends in the construction of systems of various types of maintenance (condition-based maintenance (CBM); predictive maintenance (PdM); self-maintenance and self-recovery); rather their choice in a particular situation is usually determined by the technical state of engineering at the time. Accurate information about its operating condition and prognosis of all detected trends to the side of failures or increasing risks or failures are necessary for the correct determining the optimal maintenance strategy and in general strategy of operational service of engineering object.

Thus, the life cycle of engineering relatively the types of operational service and maintenance methods is divided into several time intervals determined by the conditions of the object and prognosis. Their maintenance sets of measures are determined at each time interval from operating measures of self-maintenance to end of life management: reuse-remanufacture-recycle. Choice of various types of maintenance is more effective at accurate determining the state of engineering at the failure progression timeline. Account of optimization problem of operating costs throughout the life cycle of engineering in many respects is determined by exact definition of engineering positions on the failure progression timeline and the rate of evolution to failure state of engineering at particular time interval.

Thus, different types of monitoring for engineering are required for the construction of systems of local RUL assessments. The considered problems of the state and RUL assessments are solved by integration of remote computing PHM clusters and on-board diagnostic systems of engineering object.

On the basis of what has been said in the first chapter the formalized definition of failure progression timeline based on the principle of degeneration of failure predictors of each previous level given. The basic principle allows to construct a system of hierarchical algorithms to identify predictors of failures at every level of the hierarchy of failure progression timeline. It is also the basic features of each hierarchical level are described in the chapter.

Next chapter demonstrates methods and algorithms of some hierarchical levels that are important for the further understanding of the unification principles and the functioning of the cloud cluster with recognizing on-board automata of mobile engineering facility operating in real time. The third chapter describes some of the recent experimental results demonstrating the basic position of the hierarchical model.

The fourth chapter is devoted to a discussion of the learning principles of recognizing on-board automata using the cloud cluster.

2 General Concept of Hierarchical Model

A pilot remote PHM system supported by a parallel computing resource as a service of cloud computing, supercomputers and grid systems—the remote computing cluster PHM has been developed. The basic attention was given to construction of effective schemes of PHM monitoring, development and testing of basic prognosis models and their classification based on the characteristics of the stochastic properties of the observed signal. This takes into account the following factors of monitoring objects:

1. availability of traditional diagnostic and control systems on the monitoring objects (on-board diagnostic system);
2. availability of computing resources in on-board diagnostic system, in mobile devices and gadgets, navigation systems, etc., as well as the possibility of remote downloading of computing applications for this mobile platform;
3. availability of opportunities of telemetric data transmission to the remote PHM cluster.

In particular, methods of signal processing and the complex hierarchical prognosis model (CH & P) and its algorithms are described. The present heat is devoted to describing the general structure of remote computing cluster, the description of models and algorithms, useful first results of the pilot version of the cluster, the discussion of problems and future prospects for the development of the remote cluster.

2.1 Model

The basis of computational algorithms of PHM cluster is complex hierarchical prognosis model CH & P, for the first time stated in paper [1]. As an example, in the papers the authors consider internal combustion engine failures, revealed based on analysis of vibration engine body or certain mechanical car and track parts: gearbox and transmission, bearings, brakes, and so on. In view of the generality of CH & P prognosis model, its methods also spread to all technical objects of high complexity: generators, wind generators, turbines, etc. Recently, these methods are adapted to PHM applications in medicine, in particular, on expansion of prognostic capability of Holter heart rate monitoring systems.

Methods of diagnosis as well as prognosis methods are based on the identification of failure sings, failures detection and determination of time or speed development of signs. For example, in the diagnosis of rotating equipment

statistical characteristics of the vibration signal, peak factor, Kurtosis factor, etc. are signs of failures. However, the following questions concerning essence of PHM methods are relevant:

- How to describe the time evolution of failure signs? This question is equivalent to estimate of RUL.
- Whether the set of signs is full, i.e. whether other signs exist? And in this question it is necessary to prove or that other signs do not exist, or to show full system of signs.
- Is there some kind of order on the set of signs of failure how and what principles determine early signs from later?

Cited here research and the results of the pilot version show that set of failure signs derived from the observed signal of sensors has a hierarchical structure as shown on Fig. 1.

The hierarchical structure of set of failure signs is the basis of CH & P model and reflects the qualitative change of trajectories of some very complex and chaotic dynamical system describing the process of the functioning of mechanism. In most cases, the explicit form of the system is unknown for reasons of complexity and the impossibility to control all of its variables.

The following principles reflect the hierarchical structure of set of failure signs and are the basis for constructing CH & P model.

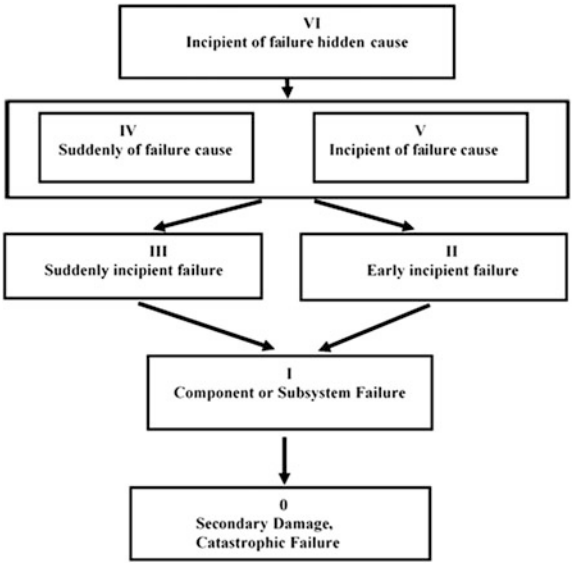


Fig. 1 Failure progression timeline and levels of hierarchy

2.2 First Level

The first level corresponds to the transition (I–0), Fig. 1 and is determined by a probability distribution function of the observed value of PDF or set of PDF wavelet coefficients of the observed value at each fixed time or rotation angle of the shaft of rotational mechanism. Signs of failures are characterized by numerical values of the moments of PDF, and their rational expressions Kurtosis factor, Hurst index, etc. Prognosis and estimate of RUL on this level is reduced to calculating time of reaching of the boundaries of states by the system (0) on Fig. 1. As a model for determining the evolution equations is used in calculating the RUL vector model of random walks of the finite segments of wavelet coefficients of the observed signal and the representation of the probability distribution function of the transition on the set of valid values of segments in the form of Feynman integrals. In this case, the evolution equations are defined in the form of the Fokker–Planck equation, the Schrodinger equation, the diffusion equation, etc. with further transition to the equations of moments. Methods for obtaining evolution equations are presented by the authors [1, 2].

2.3 Second Level

The next level is split into two sub-levels, transitions (II–I, III–I), each of which is defined as follows.

The level (II–I, III–I) appears in the case when the process becomes stationary with discrete time or its continuous variant. In this case, all of the previous signs are degenerate because does not change at mechanism operation, therefore, the time evolution of and all the characteristics in the form of combinations of moments PDF are absent.

The transition to the vector process on the level (II–I) on Fig. 1 with subsequent presentation of probability of the transition function as a Feynman integral also leads to evolution equations, but under the condition that probability distribution function (PDF) is preserved.

This condition very complicates model, however, this complication makes the RUL estimate as a time of the transition from the border (II) to the border (I) of more accurate. Sub-level (III–I) is introduced to account for the following circumstances. Under condition of preservation the fast of PDF changes in PDF, characterized by a sudden transition from one stationary PDF to the other stationary PDF, for example, the transition from single-modal PDF to bimodal, etc., are possible. These processes are simulated on the basis of the bifurcation theory and catastrophe theory. The formalized mathematical model of the catastrophe theory is contained in work [3].

In this case, monitoring of PDF parameters, which in the process of slow evolution can get in the neighbourhood of the bifurcation set is needed. The waveform

of PDF quickly is reconstructed into another. In the simplest case, it is enough to determine only two parameters of polynomial approximation of the PDF. For assessing RUL it is necessary represent the coefficients of polynomial approximation in the form of sequence of random variables. The theoretical part of the catastrophe model is developed by the authors based on the theory of noise-induced transitions [4].

2.4 Third Level

The next level and the corresponding signs are realized when the random walk of segment happens in the admissible domain, and the PDF of the transition probabilities do not depend on time.

In the cited here works of the authors at this level, estimates and prognosis were limited to estimates of Kolmogorov complexity. More detailed analysis of the results of the pilot version of PHM cluster showed that this level is split into two sublevels. Both sub-levels are characterized by of the stationary of transition probabilities. Sublevel of transitions (IV, V–III, II), as the sub-layer (III–I), corresponds to fast transitions with the change of PDF probabilities of transitions between different stationary PDF probabilities of transitions.

As well as in the previous case, catastrophe theory is used here as a model that evaluates the RUL. Sublevel of transitions (V–III, II) defined as a transition to the boundary (II, I), defined as a transition to the boundary from which the time evolution of the PDF of the transition probabilities or following stepwise changes begin (III, II–I).

RUL prognosis and estimate at this stage is reduced to the estimates Kolmogorov complexity, relative Kolmogorov complexity of all time series of vector processes defined by the wavelet coefficients of the observed signal. Details of the Kolmogorov complexity at the level (V) presented in [5]. And finally, the earliest level of the prognosis defined for today in the framework of CH&P models is determined by the transition from vector processes to processes on the orbits of the degeneracy groups of the previous level. As an example, a model of random walk in the degeneracy space of PDF transition probabilities of Gaussian type is considered. In this case, the primary time series of data of wavelet coefficients of the observed signal convert taking pairwise differences and projections on the degeneracy space, K-decomposition.

2.5 Fourth Level

Further analysis uses a model of symbolic dynamics, topological dynamics described in the works (the theory of entropy, symbolic dynamics). Well as theory of Feynman path integral on smooth manifolds is used. It should be noted that the

level of D is indicated only by the authors. Strong reasons to consider the transition to the processes on the degeneracy spaces as the system of the new class of levels splittable into multiple sub-levels have appeared now.

The transition to stochastic processes on the degeneration space of the PDF for the transition probabilities, in particular, on homogeneous spaces of degeneration space of PDF enters into the PHM practice so-called no amplitude signs.

No amplitude signs or points on the degeneration space characterize existing regularity no amplitude nature in the observed signal or its wavelet coefficients.

Representation of the transition probabilities in the form of Feynman path integral, transferred to degeneracy space, allows to calculate the RUL, i.e. the time required to achieve the boundaries of the critical condition in which changes of regularities take place in the observed signal. Change of regularities are expressed in terms Kolmogorov complexity and are evaluated using the entropy and fractal characteristics of the observed signals.

However, the transition of the system from one level to another happens not only a continuous manner, well as sudden rapid transitions to the boundary of fault are possible Moreover, if the initial distribution was one-modal, catastrophe theory describes the generation of bimodal distribution, and hence the appearance of extraneous process in operating mechanism. In the interpretation of noise-induced transitions the described situation evidence in favor transformation of the initial distribution to the distribution of with a heavy tail [6]. The prognosis estimates varies greatly towards reduction of the lifetime of the mechanism with the appearance of such distribution. However and in this case, there is an opportunity to give a numerical estimate of RUL. For the calculation of RUL is used as well Feynman integral, but in the present embodiment the obtained evolution equations for the distribution functions are defined in terms of fractional derivatives. Methods for the solution of equations with fractional derivatives are described in [7, 8].

However, at the distribution with heavy-tailed, even in the case where the system demonstrates a stationary behavior, and evolution equation has stationary solution, an important feature of heavy-tailed distributions is revealed. In a given situation of operating mechanism, if the processes defined by the entire array of wavelet coefficients of the signal has distribution with the heavy tail, a process called Levy flights is implemented at least for some in this case. Occurring in this case jumps are easily interpreted in the analysis of vibration, i.e. this is appearance of strong and single peak loadings for the high values of the scaling index of wavelet cascade. In such cases, if sufficient steepness of the front load even in the elastic region the increase in the density of defects such as dislocations in metals and alloys of the mechanism is possible. The development of dislocation by different mechanisms, in particular, Frank-Read, leads to the appearance of regions of deterioration of the material, the generation and development of microcracks. The amplitude of the elastic precursor can be in the range of valid values, i.e. in the region of linear dependence of the stress tensor from the deformation tensor. Focus mechanism of the elastic precursor contributes to the local increase in the amplitude of the elastic precursor in several times [9]. Such growth is sufficient for overcoming threshold

values of stress when the generation of dislocations per the front of the elastic precursor begins.

The result is degradation of material. Thus, the contribution to the process of material degradation from a single event, realized in the tail of the distribution exceeds the total degradation from the events taking place in the neighborhood of the maximum probability, i.e. the maximum of the distribution function. Consequently, for the appearance of a predictor or fault sign is not necessarily change in the distribution function. The solution of evolution equations may be stationary. However, in these cases of catastrophes and in the case of distribution functions with the heavy tail the nature and properties of process means that the appearance of faults is unavoidable. In these cases, the predictors will be characterized not only by the distribution function but also by characteristics of the individual trajectories.

Thus, finally structure of the basic stages of the operation of technical object throughout the life cycle is determined by the characteristics of the trajectories of states. In turn, all the valid states on the principle of degeneration of statistical and dynamic properties, for example distribution functions, are divided into classes. The set of classes is hierarchical, that is, determines the temporal phases of object functioning. The transition of permissible states from class to class means changing degrees of degradation of mechanism, the presence and degree of development of its at first the predictors and then signs of failure. In the final version, it defines the operational the age of mechanism, its remaining life.

2.6 Details of Level II

Therefore, the above-mentioned classes are hierarchically ordered. Each class is defined by the set of characteristics or predictors (Fig. 2a), which characterize:

(a) degeneracy space of class;
(b) the entire set of possible trajectories;
(c) individual characteristics of trajectories.

The evolution of the listed characteristics are developed in following scenarios, as shown on Fig. 2b:

- Slow continuous evolution;
- Fast or jump-like evolution, abruptly transforming trajectory into a class with a lower degeneracy.

Jumplike evolution has analogies with phase transitions, and the mathematical model of phase transitions and jumplike changes in types of classes or trajectories coincide.

Therefore, an important PHM task is to describe the class boundaries, their geometry, for example, the fractal characteristics, the topology of boundaries.

Because in addition to prognosis on the basis of evolution equations, which is possible under certain process conditions, such as: stationary, non-stationary,

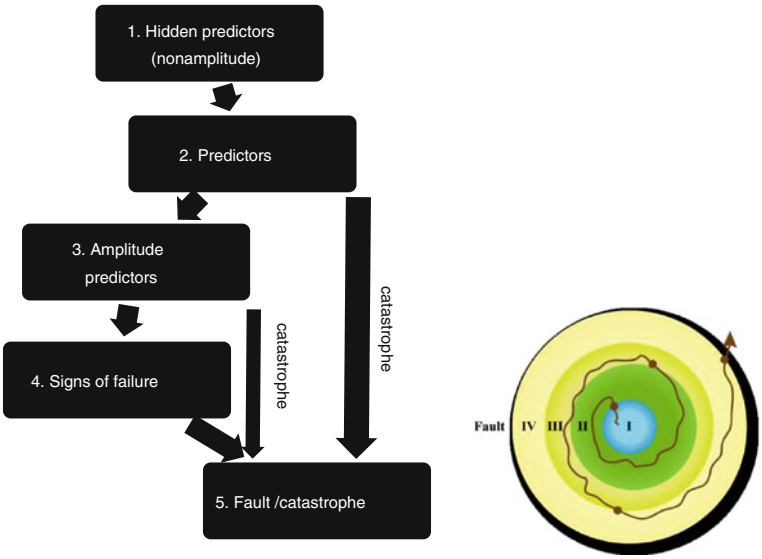


Fig. 2 **a** hierarchy of predictors characterizing classes; **b** symbolic representation of the classes and the trajectories of state

ergodicity, the smoothness conditions imposed on the coefficients of the evolution equations, etc. Besides it is necessary still to predict a temporal moment and signs at which change in the characteristics of the process, the trajectory occurs. Namely changing the type of the process generated by the change in the degree of degeneration characterizes the boundary of class.

Since in the present moment the boundaries of the classes are poorly investigated, because the commercial version of the PHM cluster creating individual chronological database is needed for their study, then during the evolution of characteristics of the trajectory, which are in fact the set of predictors, more prolonged periodic monitoring is required in real time mode. At the same time, the periodicity of such monitoring becomes more frequent at a greater distance of the predictors from the initial state. It is this more frequent procedure of periodic monitoring will affect the cost of the monitoring. A natural criterion for the effectiveness of monitoring is certainly reducing the total costs of monitoring and maintenance relative to expenses for traditional forms of maintenance without the PHM monitoring. However, during the collection of chronological database the remote monitoring cost will decrease significantly. Then the main cost reduction will be also due to the transfer of the part of functions of monitoring from the remote PHM cluster to on-board recognizing automata. Under the on-board recognizing automata here and further are meat algorithms in the form of neural networks, hidden Markov chains, Wiener nonlinear circuits, Boolean automata located on a small computing resource of on-board diagnostic systems.

Considered PHM tasks directly are related to the change of paradigms of maintenance service. The rationale for this thesis are the following arguments.

The trajectories of states of complex technical objects depend on the set of controlled parameters on the one hand and herein self-maintenance task is to keep the trajectory as long as possible within one of the classes. Self-recovery problems are reduced in context of PHM to next one, at the approach to the boundary of class by variation, possibly short-lived, return of the current state to the neighborhood of initial state. That is, the self- recovery of operating characteristics of complex technical object is implied. Far not always the solution of problems of self-maintenance and self-recovery is possible due to the thermodynamic, the operating restrictions, the dissipation process and irreversible processes of material degradation, but in the class of trajectories with only hidden or root cause predictors the solutions exist and they are cost effective.

In other types of maintenance, the basic problems are formulated the same way, but the types of maintenance are changed in each ordinary hierarchical class. Cost of maintenance increases in classes closer to the boundaries of fault, as the increase of degradation requires the use of more costly measures for the restoration of operating characteristics of engineering.

The optimization problem of the maintenance cost exist in classes based more closely to the boundary of fault, because partially in these classes the problem of self-maintenance and self-recovery remains.

3 Some Experiments

Diesel engines of trucks, which being in continuous operation, have been subjected to long-term monitoring. The vibration sensor signals of engine body in a neighborhood of block of cylinders have been registered in monitoring. The purpose of monitoring is to confirm or deny the algorithms described above for determining class boundaries and assess the RUL in each class. Experimental values of the vector processes formed from wavelet coefficients of the vibration sensor signal and the number of motor shaft rotations have been measured in the condition of periodic monitoring. In this case, the same type engines have been chosen with a different duration of operation, divided into three groups according to the duration of operation: a short period of operation, the average period, a long period of operation.

The experimental values of dispersion of the Euclidean norm of sets of segments of wavelet decomposition coefficient are shown on Fig. 3. The squares is the engines of a small operation period up to 2 years, triangles—the average period, point—a long period of operation. The theoretical curves for the dispersion, calculated on the basis of Feynman integral of the algorithm approximate the experimental data by method of variation of the free parameters. The presented data support the validity of the estimates RUL by hierarchical algorithms of PHM cluster.

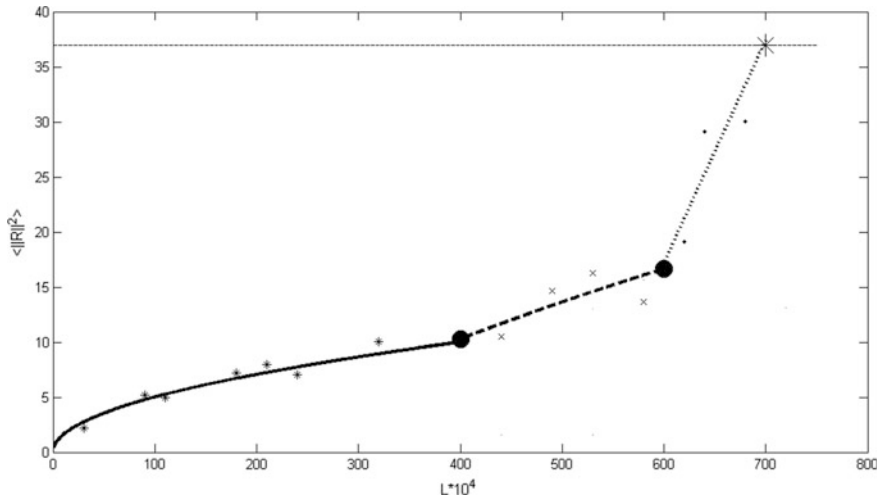


Fig. 3 The theoretical L as a function of $< \|\mathbf{R}^2\| >$ graphs and their experimental values; ⊕—component or subsystem of failure

4 Summary

Construction of the hierarchical model of failure predictors is not yet the completed problem. The hierarchy will undoubtedly expand in the process of development of computing clusters on the basis of deeper analysis of chronological databases of failure progression timeline. However, the presently available data on the structure of the predictors is quite enough for effective strategy maintenance. Reciprocal correspondence relationship between the levels of the hierarchy and the types of maintenance allows to make algorithmization of choice of the necessary maintenance in making of various constraints for example minimization of cost while maximizing the mileage with observance of the conditions of repair appropriateness. Thus, choosing the right strategy is the optimal management problem and is formalized in the language of the PHM tasks, i.e. in terms of predictors of classes, local estimates of the RUL. Therefore, optimization of maintenance procedures are a component part of the PHM. Thus, for the most cost-effective maintenance and minimizing of maintenance costs throughout the life cycle of the mechanism the need for remote monitoring PHM becomes more apparent.

References

1. Kirillov A, Kirillov S, Pecht M (2012) The calculating PHM cluster: CH&P mathematical models and algorithms of early prognosis of failure. In: Conference on Prognostics and system health management, PHM, Beijing, China. doi:10.1109/PHM.2012.6228771

2. Kirillov S, Kirillov A, Kirillova O (2011) Theoretical models and market architecture of PHM monitoring systems. In: Prognostics and system health management conference, PHM-2011, Shenzhen, 24–25 May 2011, China. doi:10.1109/PHM.2011.5939490
3. Arnold V, Varchenko A, Husein-Zade S (1982) Singularities of differentiable mappings. Classification of critical points, caustics and wave fronts. Nauka, Moscow (in Russian)
4. Horstemke W, Lefever R (1984) Noise-induced transition, theory and applications in physics, chemistry, and biology, Springer, Berlin
5. Kirillov A, Kirillov S, and Kirillova O (2011) Algorithmic method of analysis of time series data for definition of prognostic parameters of engine fault. In: 3rd international conference on advanced computer control, ICACC 2011. doi:10.1109/ICACC.2011.6016384
6. Feller W (1970) An introduction to probability theory and its applications, vol 1, 3rd edn. Wiley, New York
7. Zaslavsky G M (2002) Chaos, fractional kinetics, and anomalous transport. Physics Reports
8. Laskin N (2007) Levy flights over quantum paths. Commun Nonlinear Sci Numer Simul 12:2–18
9. Kirillov S, Klubovich V, Vagapov I (1985) Focusing of elastic pulse in solids with dislocation cluster. Izvestiya AN BSSR, №6 (in Russian)

Open Issues for Interfaces on Spare Parts Supply Chain Systems: A Content Generation Methodology

Danúbia Espíndola, Ann-Kristin Cordes, Carlos Eduardo Pereira, Bernd Hellingrath, Bernardo Silva, Átila Weis, Marcos Zuccolotto, Silvia Botelho and Nelson Duarte

Abstract This paper proposes a methodology for content generation to visualization interfaces applied spare parts supply chain systems. Case studies in the context of transportation planning for oil and gas industry will be investigated in order to validate the methodology and to analyze the results. The goal is to provide a methodology to manage and integrate the information from different systems in order to present effective data in visualization interfaces of supply chain systems.

Keywords Supply chain visualization · Content generation · Data management

D. Espíndola (✉) · C.E. Pereira · B. Silva · Á. Weis · M. Zuccolotto · S. Botelho · N. Duarte
Federal University of Rio Grande do Sul, Porto Alegre, Brazil
e-mail: dmtdbe@furg.br

C.E. Pereira
e-mail: cpereira@ece.ufrgs.br

B. Silva
e-mail: diou.bernardo@gmail.com

Á. Weis
e-mail: atilaweis@hotmail.com

M. Zuccolotto
e-mail: marcos.zuccolotto@gmail.com

S. Botelho
e-mail: silviacb@furg.br

N. Duarte
e-mail: dmtnldf@furg.br

A.-K. Cordes · B. Hellingrath
Westfälische Wilhelms-Universität Münster, Münster, Germany
e-mail: cordes@ercis.uni-muenster.de

B. Hellingrath
e-mail: bernd.hellingrath@wi.uni-muenster.de

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_129

1 Introduction

The spare parts supply chain management of large industries represents processes complex whose decision-making is not a trivial activity. Visualization methods that enable access and management of information involved in these systems through advanced interfaces can provide significant gains in terms of cost and time for industry. The information access at right time and in the correct location of the plant will allow precise actions by operators and faster decisions by managers [1].

In this sense, the use of mixed/virtual techniques can aid the interfaces from supply chain systems. However, for the effective use of mixed/virtual reality as a possibility of visualization for these systems is necessary that interfaces are integrated supply chain systems. On the other hand, a huge amount of information and different data models from several systems represents a challenge in terms of extraction and definition data, i.e. the definition, about which data should be presented in visualization interface, is not simple [2].

The real data mixed in virtual environments as well as the virtual data in real scenes can improve the visualization of some processes reducing time and costs. In this study, a methodology for virtual content generating aims to describe attributes necessary to visualize data regarding specification and location of part. Associated with the static data of part, dynamic data concerning current state of component (part) should be available at any moment. So the methodology should be able to associate different data formats and contents to parties of interest in the supply chain through integration of different systems.

The dynamic content generation for visualization interfaces of supply chain systems aims:

- to provide the 3D scenario of supply chain;
- to provide the access and the visualization in real time regarding the data and material flow;
- to represent virtually different actors and processes within the supply chain;
- to monitor the inventory;
- to identify previously (not at failure time) the location where the replacement should occur.

The integrating of dynamic content regarding part depends on the use of systems for information acquisition in real time, i.e. sensoring and instrumentation of components through sensors. In addition in order to associate knowledge to these information is necessary techniques of signals processing for prediction and diagnosis of part behavior. In this sense, the methodology will propose the data integration from intelligent maintenance systems associated with supply chain systems and virtual data related the part in question.

Following the next section describes the conceptual background regarding mixed reality systems and aspects of the spare parts supply chain where these techniques can be applied. Related works about simulation and visualization

interfaces for manufacturing processes are presented. The Sect. 3 presents the content generation methodology for visualization interface of the supply chain. Section 4 discusses the preliminary results and indicates future work.

2 Advanced Visualization and Supply Chain Systems

Currently visualization devices and advances in graphics technologies enable new ways of interaction and visualization in many different contexts. For manufacturing industry the use of augmented reality is investigated mainly for maintenance and inspection activities in machinery [3]. However, several applications are discussed in the context of mixed/virtual reality applied to industry. This section aims to present a conceptual overview about both areas of converging themes in search of their potentiality.

2.1 Mixed Reality

The mixed reality (MR) can be classified as: augmented reality (AR) and augmented virtuality (AV). The Augmented Reality is the predominance of real environment, when some virtual objects are brought to the real scene to assist in visualization. The augmented virtuality, instead, the virtual environment is predominant and real objects are brought in order to interact with the virtual scene. The AR is defined by Kirner as "the insertion of virtual objects in the physical environment, shown to the user in real time, with the support of some technological device, using the interface of the real environment, adapted to visualize and manipulate real and virtual objects" [4].

The augmented virtuality enables to visualize the virtual production chain with real data overlapped. The real data can come from sensors/RFID integrated to plant components (parts). This integration between real data and virtual scenario can provide the status of parts in real time. The augmented reality allows present virtual information on component's local in order to provide the information in the place and time needed; such AR technique is suitable for managers of monitoring and inspection of machinery/equipment in plant. The use of mobile devices integrated with augmented reality techniques are very useful for applications where the user needs walking in plant to providing feedback about the components situation. Maintenance activities are also potential applications of augmented reality, once the operator has a predominant vision of the real component with the support of virtual information [5].

Mixed scenes to data visualization from supply chain processes can be strategic in different contexts such as: visualization of parts moving of the supply chain; visualization of the components to be replaced or defendants by plant; visualization component/machine specification on equipment local (Fig. 1).

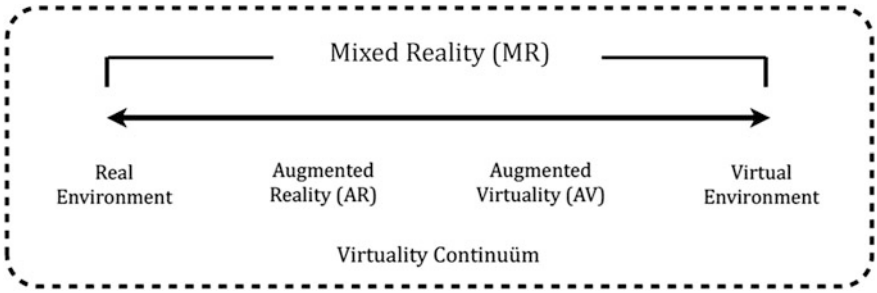


Fig. 1 Mixed reality (Milgram 1994)

Thus, in order to identify how the visualization can improve the processes of the supply chain was decided to test:

(1) the augmented virtuality integrated intelligent maintenance system—real data from intelligent maintenance system superimposed on virtual part;
(2) the augmented reality—to present data from supply chain system on real part. Other future investigation will approach the inventory management and transportation logistics.

With augmented reality use is possible to indicate the part location in stock and simulate parts placement for a more efficient logistics. The AR use can also provide information regards which parts are being loaded into the vehicle through virtual information superimposed on the vehicle. The AV use allows loading visualization—with augmented virtuality is possible simulate the material and data flow in plant with updated information in real-time such as: positioning, quantity, specification and so on.

2.2 Spare Parts Supply Chain

Spare parts are products which enable the maintenance and recovery of the original state and functions of a technical system [6]. Hence spare parts management is defined "as the management, planning and control of material and information flows, processes, activities, spare parts (as well as maintenance service personal) in support of repairing and maintaining" of technical systems of the end customer in a spare parts supply chain [7]. Analyzing the spare parts supply chain from the manufacturer's perspective, the manufacturer provides spare parts as an after-sales service to satisfy end customer needs. From this perspective four stages (supplier, manufacturer, logistics service provider (LSP), distribution center, service center and the technical system located at the end customer) can be observed in a spare parts supply chain illustrated in Fig. 2.

The main challenges in providing the spare parts as an after-sales service are the high level of uncertainty and the absence of just-in time availability due to the

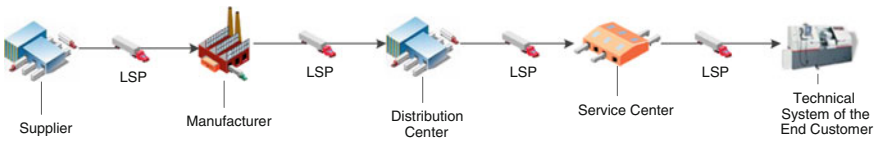


Fig. 2 Stages of a spare parts supply chain

sporadic characteristic of the spare parts demand [8]. Furthermore, the manufacturer has to deal with an extremely high variety of different kinds of spare parts and the unstable as well as fluctuating character of consumption depending on the stage of the lifecycle of the technical system. Besides all these challenges the manufacturer has to provide the spare part when it is required. Therefore, the manufacture has to organize that the required equipment, the needed spare part and the appropriately qualified service personal is in the adequate amount, at the right place and the requested time available for the replacement [9].

Each stage of the supply chain has different viewing needs. The visualization required by manufacturer may not be the same necessary for the customer. However the same view with distributed access can be interesting for different actors. In this direction future work will investigate the specialty of each demand in order to validate the methodology in case studies of different stages of supply chain.

3 Methodology

The methodology for content generating proposed in this paper identifies the following stages for the implementation of advanced visualization in systems of supply chain (Fig. 3).

(1) Modeling
(2) Tracking
(3) Data management
(4) Visualization

The next sub-section will describe each stage of methodology.

Fig. 3 The methodology proposal

3.1 Modeling

The modeling stage involves three phases: VR (virtual reality), IM (intelligent maintenance) and SC (supply chain) data modeling. The VR modeling describes the part's structure (component/equipment) and the virtual model of components. The virtual modeling also includes text information and 2D objects such as graphics and multimedia features related to real component (part). The IM modeling extracts and organizes the data generated from intelligent maintenance system about the component/part analyzed. The SC modeling will search data from supply chain system that are important for visualization.

Once defined the data model it is possible to propose a descriptive model for applications of supply chain. Through data model it is possible to simulate and to visualize the manufacturing processes. The descriptive model of the different data will provide the basis for creation of a XML relational between different models.

The TAG represents a component or part that has id attributes required for identification of material and data flow. Thus each TAG can be described by the following attributes: (i) id (identification number); (ii) position (the location in plant/stock); (iii) specification (technical specification of part); (iv) quantity (number of parts in stock).

3.2 Tracking

After the modeling stage of SC, VR and IM data, the material and information tracking can be made based on the model proposed in Fig. 4. This means, that for

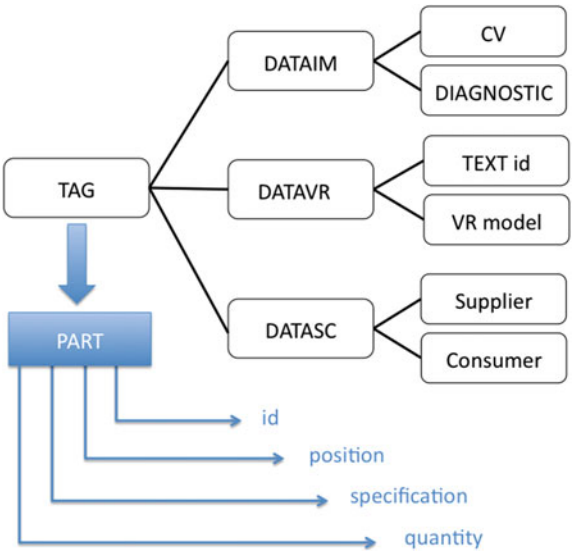


Fig. 4 The relational data model

content generating of visualization interfaces, the data described in the relational model must be extracted from the SC, IM and VR systems according to attributes defined in the model. In the short term when the part is required in the plant, the material should already be warehouse replenishment. The identification, position, specification and quantity data should be updated frequently to each part. Thus, previous planning actions for supply, production, distribution can improve the effectiveness in the long term.

Each part is identified and tracked using a tag which can be implemented by a RFID or other identification sensor. The monitoring of a large set of parties should include the management of sensor networks for monitoring the plant condition. Demand studies and evaluation of critical components of manufacturing activity should be implemented before application of advanced methods for monitoring and visualization. This means that the cost benefit to adopting tracking/visualization technologies should be analyzed before investing in the proposed solution.

3.3 Data Management

The acquisition and storage of data historical through sensors monitoring enable strategic planning of future actions related to the flow of material and data. The information management happens with data organization in specialist database where information is handled and stored in order to generate knowledge regarding the data and material behavior. The stored content is extract based on relational model described in Fig. 4. In addition, routines for information extracting and superimpose on the interface are also activities of this phase of the methodology.

The integration of different data models on the same TAG allows the data relationship from different systems to improve the visualization system. This way, display interfaces can bring data from other systems improving the access to information. The tested systems in integration for validation of the methodology were: watchdog agent (intelligent maintenance system) [10] and OpenGL [11] and ARToolKit [12] libraries as augmented reality/virtuality system. Simulating generated the data of supply chain system, i.e. it was not tested integration with the SC system, the data were manually inserted.

3.4 Visualization

The visualization stage consists of content presentation for user and providing possibilities of interaction and knowledge assimilation from the display screens. It is necessary that the interface indicate to the operator the best visualization option for each type of application. This study aims to investigate the use of two types: augmented reality and augmented virtuality. These interfaces mix real and virtual content in order to assist access to information and improve the visualization process.

Two main techniques are used to mix virtual and real components: marker use and AR markless. To position virtual elements on the real scene it can use markers, i.e. labels that are placed in the scene that when tracked by image processing algorithms, the virtual object is overlaid on the marker. Techniques without markers use image's patterns for overlay of virtual elements. The advantage in the use of markers is the lowest processing cost. For these tests, it was used the ARToolkit library to overlay and tracking real scenes. In augmented virtuality the real data were associated with object moving through rendering routines of the OpenGL library.

Discussion of preliminary tests and future work are described in the next section.

4 Discussions and Future Work

The use of advanced visualization techniques for information representation should be adopted when the gain in terms of processes time to overcome the investment for use of these technologies. Although in terms of economic value the deploying advanced interfaces is reasonable, in terms of time for development the cost is significant [13]. Thus, expert users must devote time to design and develop visualization solutions whose results are advantageous in terms of cost and time.

Thus, this study aims to propose a generic methodology for developing and implementing visualization interfaces whose contents are autonomous and intelligent. That is, the methodology should provide a means of dynamic content generating according to information provided by the supply chain and intelligent maintenance systems. Thus, the contents are not static and support automatic interaction and communication with specialist databases.

Two case studies were tested for validation of the methodology proposed and they are being discussed in order to improve the visualization processes. The first tested using augmented reality for identifying data related part (in this case specification and location of vehicle). And the second shows the use of augmented virtuality as a way to view real-time data flow and its movement in plant. The real data is represented in red on right side of Fig. 5.

Fig. 5 The mobile augmented reality (*left*). The Augmented virtuality (*right*)

In augmented reality application was used mobile device for testing, in augmented virtuality desktop device was used. Several challenges arise from applications of mobile devices. The main one is communication. The data transfer using wireless network is not always effective when the size and complexity of the models are great. Also the definition of the attributes of the relational model must still be refined during the case studies. The information extraction for SC system has been implemented with hypothetical data. Only the integration with intelligent maintenance system and VR system was tested. Finally case studies on plant should be tested in order to obtain definitive conclusions regarding the proposal.

Acknowledgments The research leading to these results has received funding from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) in cooperation project, between Brazil and Germany—BRAGECRIM, named Integrating Intelligent Maintenance Systems and Spare Parts Supply Chains (I2MS2C) in the context of visualization. This paper has been inspired by the activity of FURG, WWU, UFSC and UFRGS groups in collaboration with Petrobras and Coester industries.

References

1. Espíndola D, Frazzon E, Hellingrath B, Pereira C (2012) Integrating intelligent maintenance systems and spare parts supply chains. In: Information control problems in manufacturing, 2012, BUCHAREST. 14th IFAC symposium on information control problems in manufacturing, vol 14
2. Espíndola D, Fumagalli L, Garetti M, Pereira C, Botelho S, Henriques R (2013) A model-based approach for data integration to improve maintenance management by mixed reality. J Comput Ind (Elsevier) 64:376–391
3. Friedrich W, Jahn D, Schmidt L (2002) ARVIKA augmented reality for development, production, and service. In: Proceedings of the IEEE/ACM international symposium on mixed and augmented reality, ISMAR 2002, pp 1–13
4. Kirner C, Siscoutto RA (2007) Fundamentos de realidade virtual e aumentada. In Livro: Realidade Virtual e Aumentada: Conceitos, Projeto e Aplicações. Porto Alegre: SBC, Sociedade Brasileira de Computação, pp 2–21
5. Henderson S, Feiner S (2011) Exploring the benefits of augmented reality documentation for maintenance and repair. IEEE Trans Vis Comput Graph 17(10):1355–1368
6. Hellingrath B, Küppers P (2011a) Model-driven development of multi-agent based collaborative planning concepts for heterarchical supply chains. In: Maríık V, Vrba P, Leitão P (eds) Holonic and multi-agent systems for manufacturing. Springer, Berlin
7. Kutanoglu E, Mahajan M (2009) An inventory sharing and allocation method for a multi-location service parts logistics network with time-based service levels. Eur J Oper Res 194 (3):728–742
8. Hellingrath B, Küppers P (2011b) Multi-agent based evaluation of collaborative planning concepts in heterarchical supply chains. In: Sucky E, Asdecker B, Dobhan A, Haas S, Wiese J (eds) Logistikmanagement: Herausforderungen, Chancen & Lösungen; Band III, Bamberg
9. Loukmidis G, Luczak H (2006) Lebenszyklusorientierte Planungsstrategien für den Ersatzteilbedarf. In: Barkawi K, Baader A, Montanus S (eds.) Erfolgreich mit After Sales Services. Geschäftsstrategien für Servicemanagement und Ersatzteillogistik, Berlin, pp 251–270

10. Djurdjanovic D, Lee J, Ni J (2003) Watchdog Agent, an infotronics-based prognostics approach for product performance degradation assessment and prediction. Adv Eng Inform (Elsevier) 17(3–4):109–25
11. OPENGL: Open Graphics Library (1997) [S.l.]: Silicon graphics interface. Disponível em: http://www.opengl.org/about/overview/. Acesso em: 13 june 2009
12. Kato H, Billinghurst M, Poupyrev I (2000) Artoolkit version 2.33 [s.1:s.n.], Manual de software. Disponível em: http://www.tinmith.net/lca2004/artoolkit/artoolkit2.33doc.pdf. Acesso em: 20 mar. 2011
13. Porcelli I, Rapaccini M, Espíndola D, Pereira C (2013) Innovating product-service systems through augmented reality: a selection model. In: Shimomura Y, Kimita K (eds) The philosopher's stone for sustainability. Springer, Berlin, pp 137–142

Research on Wear Behavior Analysis, Modeling and Simulation of the Integrated Control Valve

Fan Kejia, Xiao Ying and Kang rui

Abstract Integrated control valve in the aircraft control system has an important role to play in, spool valve, apartments and will affect a variety of performance indicators in integrated control valve wear, is one of the major failure mechanism. Based on integrated control valve structure and working principle, analysis of spool valve sleeve wear failure sensitive parts and structural parameters affected by wear. Using AMESIM software to model integrated control valve through modeling and simulation of spool valve sleeve due to wear and changes of structural parameters on performance index of integrated control valve. Finally, is worn to physical failure model, based on deviation of product performance thresholds, calculated integration time control valve failure, to propose an integrated control valve service life prediction method.

1 Foreword

Wear and tear is a typical failure mechanism of hydraulic products, current material or the wear behavior of many individual components, but for multiple components for wear life of a relatively small number of products [1], and some of the traditional methods of error is large, for example: the weakest links theory. Due to product performance while decreasing by more than one component wear and failure, so the actual life of a product is always less than the life of the individual components [2], so typically the weakest links theory is inaccurate in calculating the product life. Airborne integrated control valve, this paper studies, using AMESIM simulation software based on combining has worn physical failure model, proposed a new method for calculating life expectancy.

F. Kejia (✉) · X. Ying · K. rui
School of Reliability and Systems Engineering, Beijing University of Aeronautics and Astronautics, Beijing, China
e-mail: Fankejia18@gmail.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_130

2 Operating Principle of Integrated Control Valve

The integrated control valves are key components for the airborne aileron actuator. It is for supplying oil to the oil cavity of actuator to drive aircraft aileron deflection. Figure 1 shows the structure diagram of integrated control valve.

The integrated control valve is mainly consists of four functional valves (from top to bottom in Fig. 1 are: switching valve, communication valve, select valves, damping valve),a flat valve, a electro-hydraulic servo valve and a solenoid valve.

When flying by wire, followed by a signal turning on the solenoid valve, high pressure oil inject in the control cavity of damping valve directly, and supply oil to the other functional valves through the solenoid valves to make functional valves open.

While the high pressure oil passing though the switching valve through another channel to supply the electric liquid servo valve with control oil and high pressure oil. The electric liquid servo valve controlled by electrical signals, selects oil supply to left cavity or right cavity of the actuator.

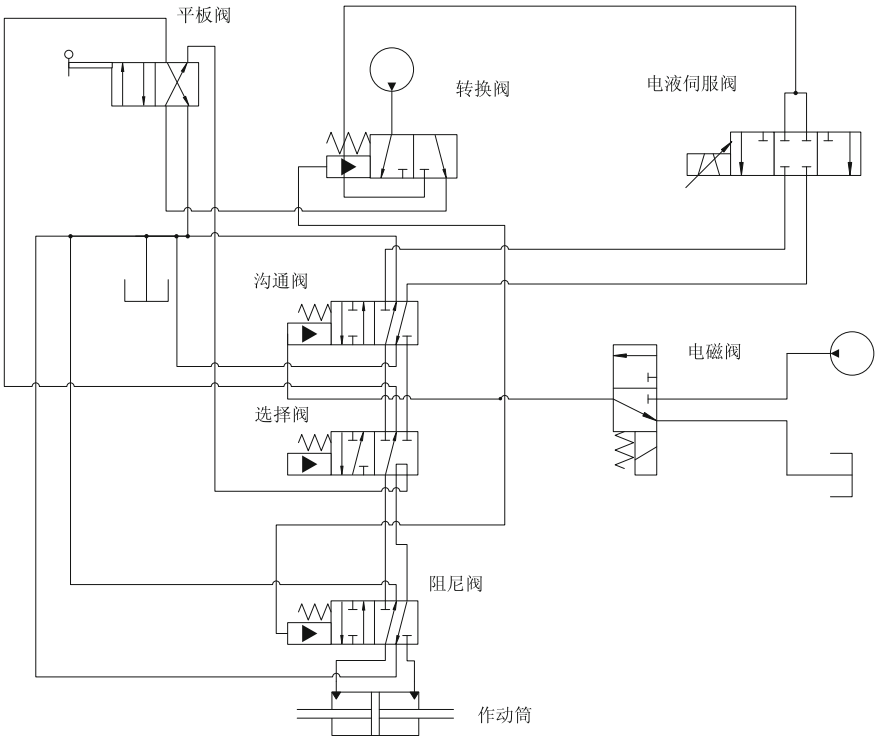


Fig. 1 Structure diagram of integrated control valve

When mechanical modal, the high pressure oil controlled by switching valve supply to flat valve, the flat valve accept mechanical operation signal, select the actuator flow to left cavity or right cavity of the actuator to drive the actuator.

3 Abrasion Behavior Analysis Principles and Processes

3.1 Determine the Wear Failure Mechanism Model

In General, according to wear mechanism and the function mode between the material and material of the friction pairs, the main types can be divided into abrasive wear, adhere wear, erosion wear, fatigue wear, corrosion wear and fretting wear [3]. Wear mechanism of the product is very important, chose the appropriate mathematical model according to the wear mechanism.

This article chooses the adhesive wear which a typical representative of the slide valve mechanism as object of study, and selects Archard Model as source model to calculate the wear life. Wear may cause the slide valve leaked, slide valve leaks are divided into external leakage and internal leakage. External leakage means oil from the internal components or tubing interface to an external leak. Internal leakage is defined as inside the components a small amount of oil due to clearance or wear flow from high pressure cavity to low-pressure Chamber. The leakage will have an effect on the output force, rapidity, and other performance parameters of the actuator.

3.2 Determine System Sensitivity Parameters

For integrated control valves in the structural parameters sensitivity analysis, identifying which have an impact on performance, identified as sensitive parameter. Sensitive parameters may be the structural parameters of the product, also may be environmental parameters and so on. Sensitivity analysis can be divided into local and global sensitivity analysis [4]. Local sensitivity analysis tests only a single parameter on the performance impact. Global sensitivity analysis tests multiple parameters of the influence on product performance. In this method, The different between local sensitivity and global sensitivity analysis only lies in calculating quantity size. Sensitive parameters need to be a combination of methods, such as through literature research, similar product experience, and the use of computer simulation and model analysis, find out one or a few sensitive parameters. It identifies four potentially sensitive parameter of the integrated control valve: spool diameter, valve inner diameter, valve clearance between core and pocket, the contact length of valve spool and pocket. Sensitivity analysis is carried out on the structural parameters of the four, finally determine the valve clearance between

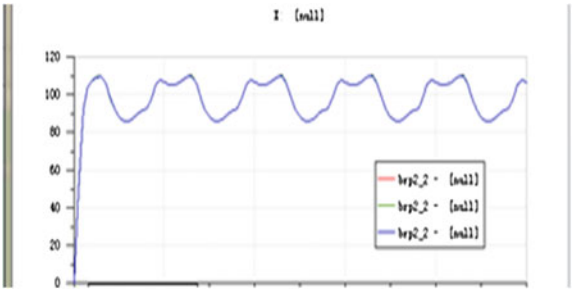


Fig. 2 Spool diameter impact on production system performance

spool and pocket as sensitive parameters. Figures 2, 3 and 4 show different structural parameters impact on product performance graphs. Clearly we can indicate that valve clearance on system performance impact greater than other parameters.

It's to be noted that, in practice, the structural parameters have an impact on system performance more or less, choices of the sensitive parameter should be based on the actual needs, so as to reduce unnecessary computation.

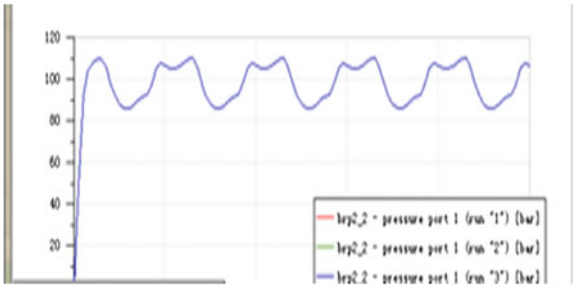


Fig. 3 The contact length of valve spool and pocket impact on production system performance

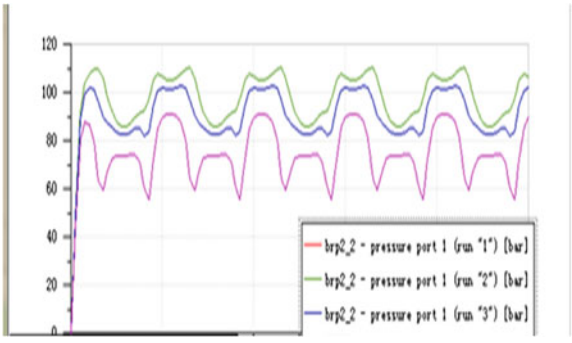


Fig. 4 Valve clearance between spool and pocket impact on production system performance

3.3 Determines the Product System Failure Criteria and Sensitive Parameter Threshold

Before determine sensitive parameter thresholds, first of all make sure the product system failure criteria which the performance parameter values does not satisfy the requirements. Products tend to have more than one failure, in this paper the object of study, for example, integrated control valve of the main performance parameters: actuator maximum output power, actuator maximum displacement, actuator piston movement speed and frequency response of the system. If one of the indices do not meet the requirements, namely that the whole system had lapsed, so select the first can not satisfy the required performance indicators: System frequency response, as a failure criteria. According to the design of indicators, frequency attenuation of response = 20 lg actuator displacement/input signal amplitude, its value should be less than 3 dB, Fig. 5. According to this failure criteria, we can use AMESIM simulation software model concluded that when the system frequency response 3 dB, all sensitive parameters value, this value is the threshold of sensitive parameter.

3.4 Calculates the Degradation of Life

From the above process (3) Calculated the threshold of each slide valve clearance, minus the assembling slide valve gap, we can get the spool valve set maximum allowable wear, substitute this value into the Archard Degradation model to calculate the wear life of a single valve. In the weakest link principle, the minimum degradation of the life $TF_{\min}$ is considered to be products system life. In fact, slide valves wear and the sensitive parameters affecting system performance at the same

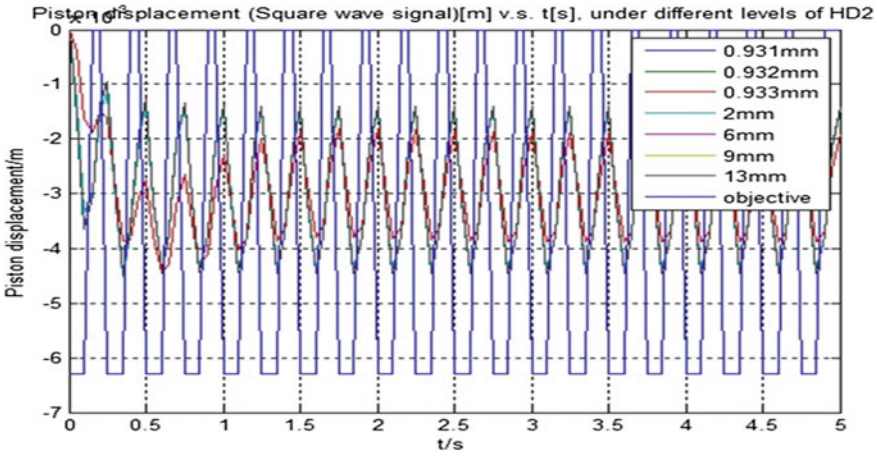


Fig. 5 The system frequency response with different clearance

time, production systems for more accurate prediction of degraded life, must be considered the spool valve sleeve of each valve clearance impact on production system performance.

4 Practical Cases

For a airborne integrated control valve, using the AMEsim Software, build the integrated control valve and its related components of physical functional model that simulates the integrated control valves under normal working status, the relationships between structural parameters and system performance [5]. To establish the foundation of looking for sensitive parameters and degradation behavior analysis for next. Figure 6 shows the fully functional model of the products building with AMEsim software.

According to the design criterion, the failure criterion of the integrated control valve is that the attenuation of system frequency response exceeds 3 dB. With the above analysis, which the clearance between the spool valve and the valve pocket is determined to be the sensitive parameter, the threshold of each sensitive parameter based on the failure criterion is listed in Table 1. Figure 7 shows the relationship between clearance and the attenuation of system frequency response of change-over valve, which is similar with that of communication valve, selector valve and orifice valve. The curve in Fig. 7 is monotone increasing. The attenuation of system frequency response reaches 3 dB when the clearance is 0.074 mm. That means the

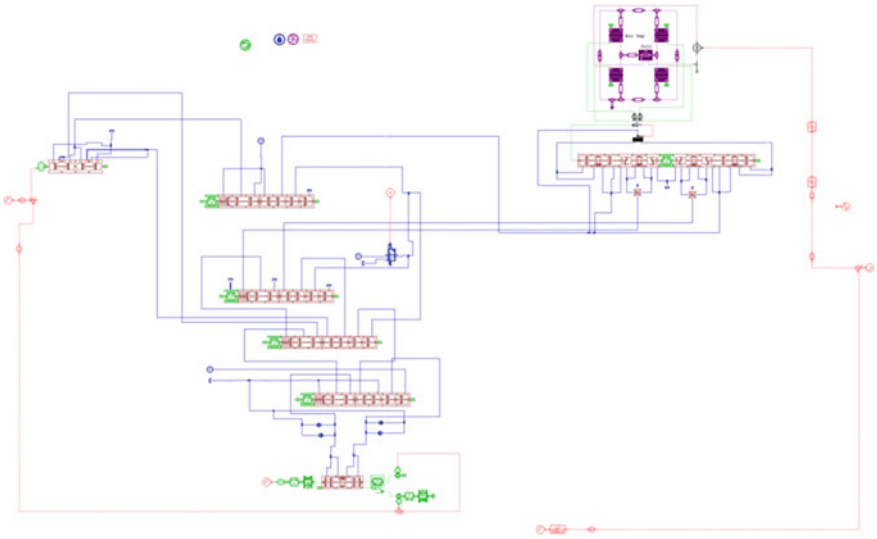


Fig. 6 System functional model of integrated control valve

Table 1 All sensitive parameters and threshold values

Unit	Sensitive parameters	Design value (mm)	Threshold (mm)
Servo valve	Spool and valve clearance	0.001	0.0538
Switching valve	Spool and valve clearance	0.007	0.074
Communication valve	Spool and valve clearance	0.007	0.0723
Select valves	Spool and valve clearance	0.007	0.072
Damping valve	Spool and valve clearance	0.007	0.075

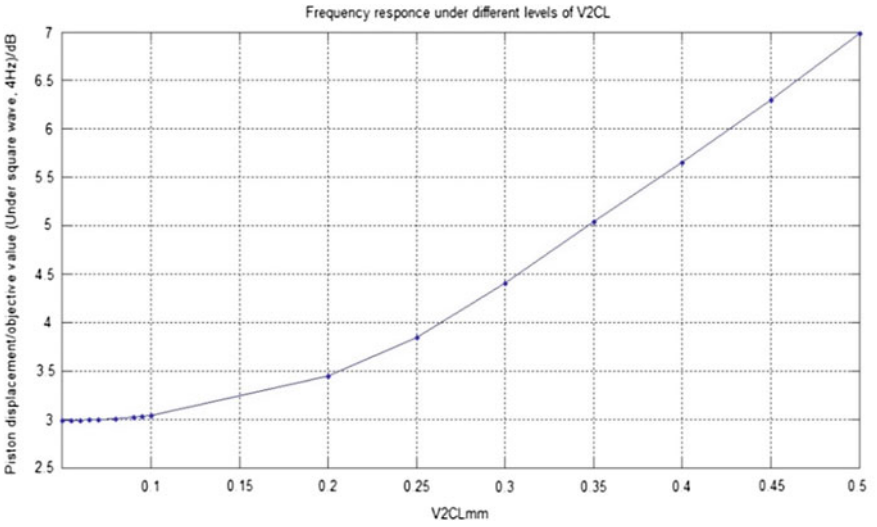


Fig. 7 Curve of relationship between system frequency response attenuator and switching valve clearance

attenuation of system frequency response decreases with the increasing of the clearance, and the system fails when the clearance exceeds 0.074 mm.

Gain spool valve set threshold minus each valve clearance gap, you can get each valve to maximum allowable wear, substitute Archard Wear and degradation of the models calculate each valve life. By taking the smallest to $TF_{\min}$ (The shortest link theory, is considered $TF_{\min}$ System for product life) [6] Further use the bisection method to calculate time of the whole integrated system control valve failure, a lower bound of $TF_{\min}$ of up to $TF_{\min}$, Calculate $t_{u1} = \left(TF_{\min} - \frac{TF_{\min}}{6}\right)/2$, Each valve wear $h = vt_{u1}$ and AMESIM Software emulation of the wear volume, system response attenuation is achieved 3 dB If attenuation system response does not exceed 3 dB And constant iterative computation, simulation; i At the time, the product system responds to decaying 3 dB, determine product system life t_{ui}.

Finally, if only one valve is involved, the life based on the shortest link is 1.68×10^7 cycles, however, when all of the valves degrade together, the life of system is reduced to 1.24×10^7 cycles.

5 Summary

The author of this paper carry out a new method of wear degradation life calculation about product constituted by many components. This method considers the internal product parts wear impact on the overall performance of the products and use computer software to simulate the functions of the product. And through the iterative method to calculate the wear life of the product. Compared with the traditional calculation method of the degradation of life, this method is more accurate and useful in engineering practice.

References

1. Gertsbackh IB, Kordonskiy KB (1969) Models of failure. Springer, New York
2. Yao Z (1988) System degradation and study of system reliability. Institute of Automation, Chinese Academy of Sciences
3. Guan C Abrasive wear failure analysis for mechanical produces. PTCA(Part: A Phys.Test.)
4. Cai Y (2008/02) Review of sensitivity analysis. J Beijing Norm Univ (Nat Sci) 44(1)
5. Wu SJ, Shao J (1999) Reliability analysis using the least-square method in nonlinear mixed-effect degradation models. Statistica Sinica 9(3):855–877
6. Mealy GH (1955) A method for synthesizing sequential counts. Bell Sys Tech J 34:1045–1080

Study on Evaluation Index System of Equipment System Transportability

Qian Wu, Lin Ma, Chaowei Wang and Longfei Yue

Abstract In recent years, with the extensive application of high-tech, more and more military equipment shows the characteristics such as multi-functional, integrated and multi-system, large-scale. And equipment system mobility requirements are also increasing in the modern war. Bad equipment system transportability will lead to late equipment delivery, reducing transportation efficiency and equipment combat performance. Whether equipment system is able to be transported timely and effectively will directly affect the quick and efficient formation of equipment system combat. So equipment system transportability should be analyzed in order to improve the operational effectiveness of equipment system. Study on equipment system transportability in domestic starts lately and there is no complete description of transport index system. In this chapter, the various factors which effects equipment systems transportability are summarized on the basis of equipment systems transportability analysis. Evaluate index system of equipment system transportability is also presented based on the construction principle, such as integrity, non-compatibility, operability. The results of this study provide a basic parameters for evaluate model of equipment system transportability. And it offer effective technical support to evaluation and design of equipment system transportability.

Q. Wu (✉) · L. Ma (✉) · C. Wang (✉) · L. Yue (✉)
School of Reliability and System Engineering, Beihang University, 100191 Beijing, China
e-mail: ustbwuqian@126.com

L. Ma
e-mail: malin@buaa.edu.cn

C. Wang
e-mail: wangcwbu@163.com

L. Yue
e-mail: bmw_13567mxz@163.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_131

1 Introduction

Transportability of weaponry is the innate ability of efficient transportation and adaptive capacity of transportation environment of weapon equipment drafted, carried or self-drove by existing or planned conveyance. It is a basic requirement that equipment system should be easily adapted to transport. The fighting capacity cannot be performed effectively if the equipment system without transportability cannot be deployed quickly. In recent years, with the application of high technology, weaponry develops toward multifunction, integration, multi-system and large-scale. Modern war is a war of high mobility, which increasingly demands higher maneuvering characteristics of the equipment system. The transportability of equipment system issues more and more obvious and plays a key factor of the formation and performance of weaponry.

As an inherent attribute, the transportability of equipment system is determined as soon as the design is completed. Therefore, the research on the transportability of equipment system is important. In order to transport the equipment effectively, safely and quickly by all kinds conveyances and enhance the combat effectiveness without compromising other performance, the transportability of the equipment system should be analyzed and evaluated in the early stages (and the whole process) of project approval, development and procurement. Compared with other equipment characteristics, such as supportability and maintainability, transportability of equipment started much later. Currently, the study of transportability of equipment system is still in the exploratory stage, both in China and overseas. In the chapter, the main factors affecting the transportation are exacted through the analysis of transport process and the evaluation parameter system is constructed based on the domestic and abroad research result of transportability of equipment system. The qualitative description of the meaning of each parameter of the system is given.

2 Work Analysis on the Transportation of Equipment System

2.1 Influencing Factors of Transportability of Equipment System

In the equipment development process, much attention is paid on the combat capability, power and automation of the equipment, and the analysis on the transportability of the equipment based on the existing transport conditions is neglected. The equipment is difficult to transport fast after deployment because of unfitted size, loading and unloading difficulties, fastening difficulties and mismatch with existing carriers, which lead to many problems such as inefficient, long transport time, high cost and poor security. The problems restricts the operational effective of equipment. So the factors influencing the transport of equipment must

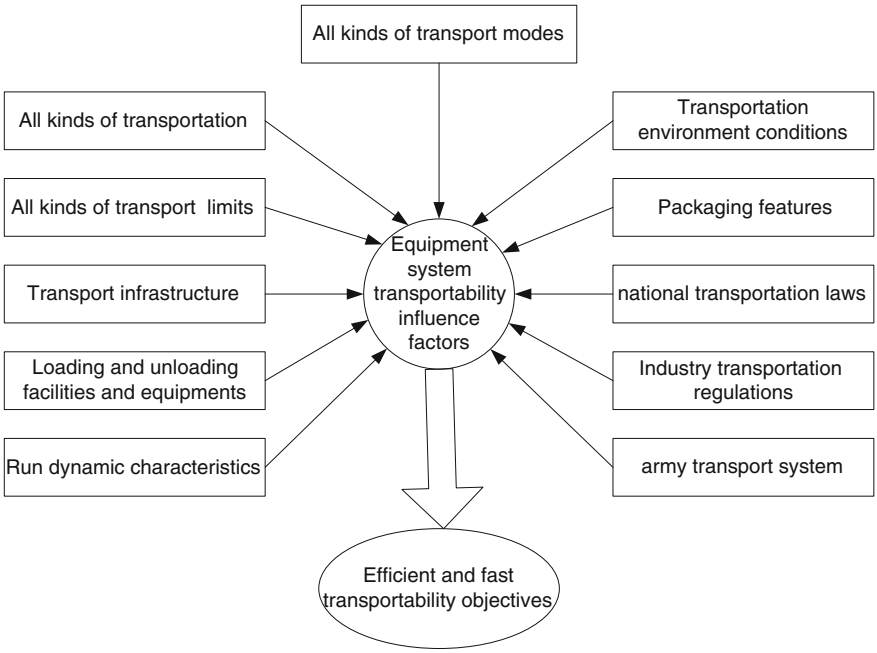


Fig. 1 Equipment system transportability influence factors

be identified and the design program should be updated to improve the transportability of the equipment and to avoid similar problems. The factors affecting the transportability are shown in Fig. 1.

2.2 Procedure Analysis of Equipment System Transportation

The transportation process of equipment system is composed of a series of activities, like packaging, loading and unloading, storage and transport. Based on the analysis of transport process affected by the factors, the transport process is shown in Fig. 2. The required packing box specification and material, outer packing, protection level, packing type of transport object on different transport methods and lines can be determined through the analysis of packaging. The loading and unloading conditions, equipment or facilities for transport equipment can be determined through the analysis of loading and unloading. The transportation resource, time and cost can be determined through the analysis of transport activities. The storage conditions and mode of the transport object can be determined through the analysis of storage. In the equipment development stage, the equipment design scheme should be adapted for the transport scheme to improve the transportability of the equipment looking forward to achieve the rapid and efficient transport objective during the using stage.

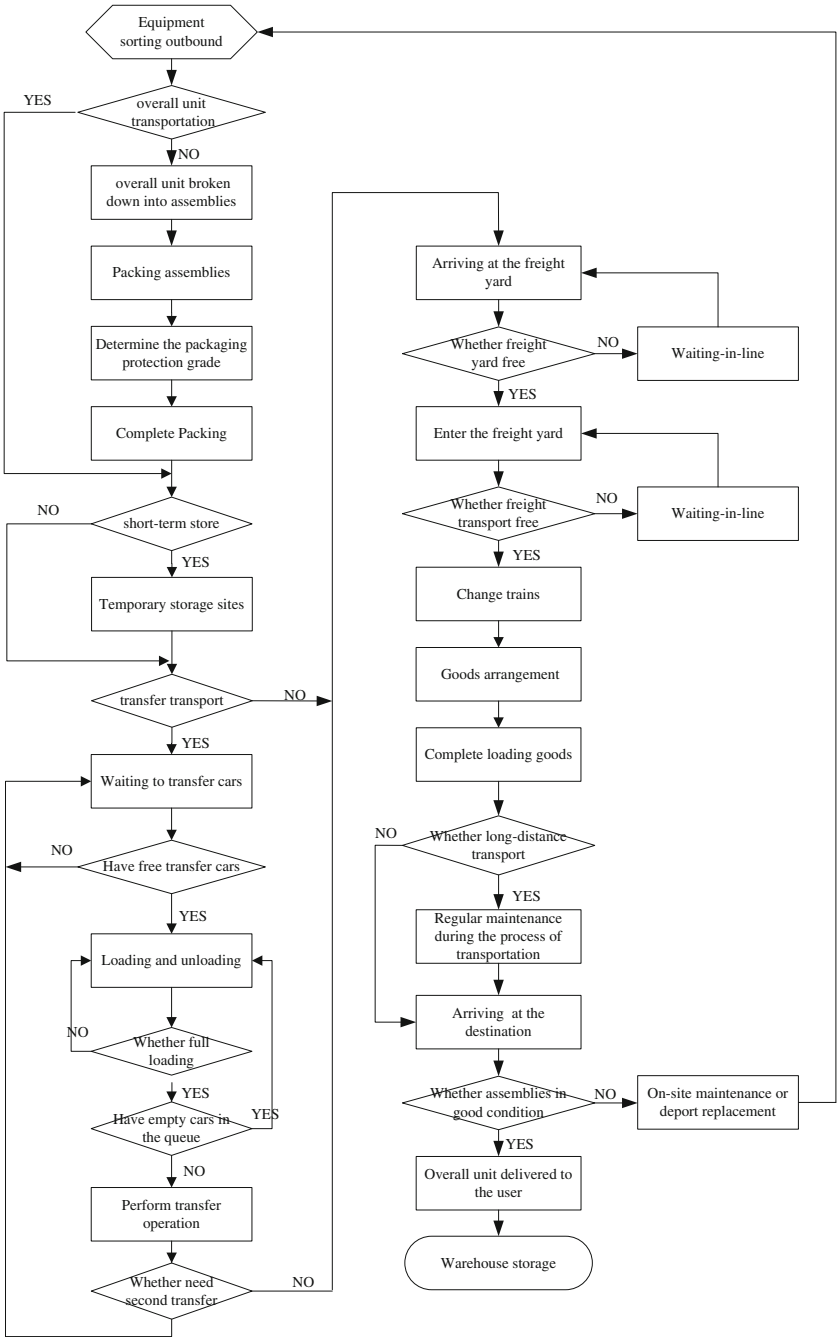


Fig. 2 The flow chart of equipment system's transport process

3 Transportability Evaluation Index System

When evaluating the transportability, the index is more complex. In order to evaluate the transportability of the equipment system comprehensively and reasonably, more than one indicator is needed. For this reason, it is necessary to conduct the evaluation index system taking the analysis of influencing factors and transport process into account.

3.1 Principles of Transportability Evaluation Index System Conduction

Logistic support of IJO stresses the systemic and holistic logistical support of the whole combat system other than single equipment. Focusing on the requirements of logistic support, the support resource, units, organizations and other elements of various armed force should make comprehensive integration, which includes navy, air force, space, and eclectic. The formation of logistic support needs to adapt to multi-factor optimization. The combination of the three major elements, as known material, energy and information, must be efficient and stable so as to provide continuous, secure, and efficient logistical support for the war.

Evaluation Objective is reflected by evaluation index. The determination of evaluation index is the key point of whole evaluation system. The selection of incorrect evaluation index would result in the failure of evaluation. Therefore, the determination and establishment of a systematic, reasonable and comprehensive index system are the premise of evaluation. The evaluation index system should be able to describe all the characteristics of an abject, so the construction of Transportability Evaluation Index System need to follow the following basic principles:

(1) Principle of Completeness. The completeness of Evaluation Index System includes two aspects: adequacy and necessity. The characteristic of any parameter selected should be described comprehensively by the sub-parameters.
(2) Principle of incompatibility. There is not compatibility between each parameters, which means one parameter cannot replace or contain one another. Some relevant parameters may reflect crossed content rather than compatible content. Relevant parameters can be contained in the index system, while the compatible parameters are not.
(3) Principle of Operability. The selected parameters should be simple and clear. On the one hand, simple parameters can be helpful to find the key point form complex information by avoiding confusion; on the other hand, simple parameters can reduce the workload for easy calculation and analysis, contributing to the evaluation easily.

(4) Principle of Standard. The standard parameters within the scope of research should be chosen for its versatility, which is convenient for the collection of date and information and the understanding of the parameters.
(5) Principle of Comparability. Index system should be able to be compared at different times to reflect and determine the operating state of support under different conditions. To expand the scope of the comparability of parameters, the existing statistical parameters should be took advantage as possible.

3.2 The Evaluation Index System of Equipment Transportability

An excellent equipment system transportability depends on the design characteristics of the system and the corresponding support resource. Therefore, according to the construction principles of the index, the Transportability Evaluation Index System is built on the basis of influential factors of equipment system transportability and the analysis of transportation process, as shown in Fig. 3.

Equipment transportation efficiency, the first level of evaluation index, is the general objective for the evaluation of equipment system transportability. The secondary parameters performance as the criterion layers of the evaluation of equipment system transportability. The transportation process of materiel system involves some functional procedures, like packaging, loading and unloading, storage and transport, which are limited by the time, cost, efficiency, quantity and versatility. In Fig. 3, the indexes representing different functional procedures are set as the secondary evaluation index; the specific evaluation parameters are set as the third level of evaluation indexes.

- Time Index

Based on the time requirement of achieving every procedure, the Time Index of equipment system transportation process can be disassembled to Packaging Time, Loading and Unloading time, Transport Time and the Waiting Time to execute every procedures. The Waiting Time is caused by the lack of transportation resources during the transport of the equipment, including the waiting time for packaging, loading and unloading, transport and storage. On equal conditions, a smaller time index is better.

- Cost Index

Similarly, according to the cost requirement of achieving all functional procedures, the Cost Index can be represented by the Equipment Cost, Expendable Cost and Staff Cost. In every procedure, there is the cost of nonexpendable fixed assets, defined as Equipment Cost, such as forklift truck, transport vehicle and warehouse. An Equipment Cost is expressed by the value: cost divided by rated service life. On equal conditions, a smaller Cost Index is better.

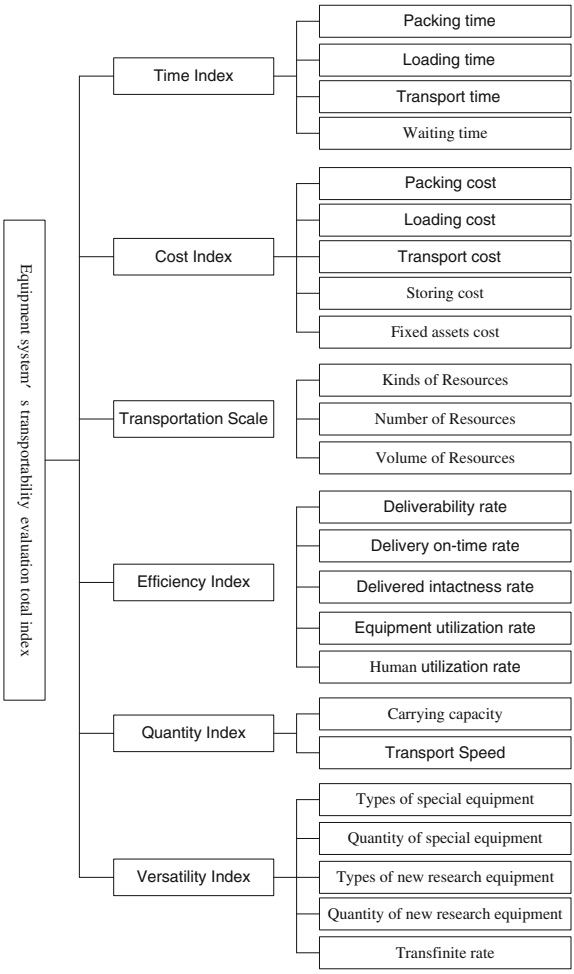


Fig. 3 The evaluation index system of equipment transportability

Index is better.

- Transportation Scale

The Transportation Scale represents the variety, quantity and volume of all the transportation resources to achieve an equipment system transportation task. On equal conditions, a larger transportation scale means a worse transportability of the equipment system.

- Efficiency Index

Efficiency Index is a symbol of the utilize situation of manpower and equipment in transportation process and the probability of equipment transported on time and in good condition under the condition of existing resource. Utilization rate of

manpower and equipment is a ratio of service time to the rated life. Deliverability rate is a ratio of the times of delivered object reached to the total delivery times. Delivery on-time rate is the ratio of the times of equipment delivered on time to the total delivery times. Delivered intactness rate is the ratio of the times of equipment reached in good condition to total delivery times. Efficiency Index is a mean value. On equal conditions, a lager Efficiency Index is better.

- Quantity Index

Quantity Index represents the adaptability of the equipment system to the transportation means. Carrying Capacity symbols the allowable maximum traffic volume of the equipment during the transportation, such as flammable, explosive materials and other dangerous object. Transport Speed is the maximum transport speed under the equipment system intact premise.

- Versatility Index

Versatility Index is used to indicate that the versatility of the transport resource needed in equipment and system transportation, described by types of special equipment, quantity of special equipment, types of new research equipment, quantity of new research equipment. Under same conditions, smaller the types and quantities of special equipment and new research equipment mean a better transport resource versatility of the equipment system.

The evaluation index system of equipment transportability is set up on the basis of the analysis of influencing factors of equipment transportability and transport process. When evaluating the transportability of equipment system, the third-level index of the bottom should be calculated firstly. Then, the third-level index can be used for the estimation of the secondary index, and repeat the process for the objective first-level index.

4 Summary

On the basis of analysis of influencing factors of equipment transportability and transport process, the Transportability Evaluation Index System of equipment system is built up by Structure Module Method with three layers: Objective Layer, Principle Layer and Bottom Evaluating Parameters Layer. All Evaluation Indexes are set from top to bottom to refine the transportability of the equipment adequately. In addition, the relevant parameters are defined and a basic algorithm is provided. In this method, the index system possesses such advantages as good integrity, high operability and low repetition. The Index System can be used for multiple assessment method, evaluation object and evaluation occasion which support for the further researching on evaluation system of equipment system transportability.

References

1. Liu ST (2005) Research on Military Equipment Transportation Engineering Theory, Method and Application B
2. Transportability Testing Procedures (2004) US Army Defense Ammunition Center, July 2002, Virtual Proving Ground, Military Traffic Management Command, 27 Aprll 2004
3. Transportability Analysis Reports Generator (1998) Military Traffic Management Command Transportation Engineering Agency, 10 Feb 1998
4. US Department of Defense (1998) MIL-STD-209 J, Interface Standard for Lifting and Tiedenm Provisions, 28 Jan 1998
5. Department of Defense (1998) MIL-STD-1366D, Interface standard for Transportability criteria, 18 Dec 1998
6. US Department of the Arm (1985) Army Regulation 70–47, Engineering for Transportability, Army Publications and Printing Command, 19 Aug 1985

Corrosion and Protection Status in Several Chinese Refineries Processing High-Acid Crude Oil

Chunlei Liang, Xuedong Chen, Yunrong Lv, Zhibin Ai and Junfeng Gao

Abstract Naphthenic acid corrosion and corrosion protection status in several Chinese refineries processing high-acid crude oil was summarized and analyzed, including material selecting, corrosion monitoring methods, corrosion case. The processing experiences indicate that the most several corroded parts when processing high-acid crude oil occurred in the vacuum column and transfer line. In one refinery, vacuum tower wall, tower packing and transfer line with SS316L (Mo ≥ 2.5 %), were corroded amazing after 3 years operation, which need to be upgraded to SS317L and with the use of corrosion inhibitor to enhance corrosion protection. In another refinery, many high temperature heat exchangers made of carbon steel in the distillation unit were found corrosion severe. In the third refinery, the atmospheric transfer line with SS321 material had a corrosion rate of 0.3 mm/year, found through the fixed-point thickness measurement, and then upgraded to SS316L. From the processing experiences, it could be concluded that selecting SS18-8, SS316L or even SS317L is necessary when processing high-acid crude oil. Corrosion inhibitor is suggested to be used in the most sever corrosion part, for example in the vacuum column and transfer line. Naphthenic acid corrosion of SS316L has an incubation stage. Corrosion is slower in the initial stage, following by a rapid development once the corrosion forming which made the surface of the metal rough. Therefore, every time a unit shutdown, corrosion inspection should be carefully conducted.

1 Introduction

According to statistics, the global high-acid crude oil (Total acid number, TAN ≥1.0 mg KOH/g) production has accounted for 10 % of global oil production in recent years. High-acid crude oil processing capacity of Sinopec, PetroChina, CNOOC increased rapidly in recent years because of the higher profits. To 2009,

C. Liang (✉) · X. Chen · Y. Lv · Z. Ai · J. Gao
Hefei General Machinery Research Institute, Hefei 230031, Anhui, China
e-mail: lchunlei@126.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_132

Sinopec had more than 20 refineries processed high-acid crude oil. High-acid crude oil processing capacity of Sinopec in 2009 exceeded 30 million tons, accounting for 21 % of total processing amount [1]. In 2009, CNOOC Huizhou refinery putted into the first set of 10 million tons distillation unit processing high-acid crude oil in China.

A variety of corrosion problems had been occurred when processing high-acid crude oil, whether in Sinopec, PetroChina or CNOOC. Equipment corrosion has become the main issue for constraining the unit's long term operation.

Damage due to naphthenic acid corrosion (NAC) was first observed in 1920s [2], and articles were published since 1950s [3]. Some achievements on corrosion mechanism and protection methods of NAC have been made now [2–6]. However, corrosion has not been completely resolved, still lacking a widely recognized relationship between NAC and various affecting factors, for example materials, sulfur content, temperature, fluid velocity [2, 5].

Selecting higher grade material is generally considered the main way to dealing with high temperature NAC. Most of the corrosion problems could be avoided using higher grade material combined with corrosion monitoring and corrosion inhibitor. Corrosion and protection status in several Chinese refineries processing high-acid crude oil was summarized and analyzed, including material selecting, corrosion monitoring methods, corrosion cases. Some problems such as corrosion of vacuum column and transfer lines were analyzed and discussed. This chapter hopes to provide helpful references on corrosion and protection for other companies processing high-acid crude oil.

2 Material Selecting

2.1 Material Selecting Guideline

Reasonable materials selection is the main approaches to dealing with high temperature NAC. Chinese companies have also done some tries and researches. China Sinopec developed the refinery materials selection guidelines in 2001 and 2002. With the processing capacity of high-acid crude oil increasing, equipment and piping damage caused by NAC have been one of the focus problems in refining industry. The material selection guideline was revised since 2008 and issued in 2012, named SH/T 3129-2012 "material selection guideline for design of equipment and piping in units processing acid crude oils". The new guideline referred the successful experiences and corrosion issues in Chinese refineries, as well as material selection in foreign refineries, also part use of the latest research results of American Petroleum Institute (API) and NACE International. The main material selection criteria of SH/T 3129 are shown in Table 1.

Table 1 Main material selection criteria of SH/T 3129

Temperature /t	Material selection of equipment	Material selection of piping
t < 240 °C	Carbon steel	Carbon steel
240 °C ≤ t < 288 °C	SS304L	v < 3 m/s, 5Cr1/2Mo or SS304L
		3 m/s ≤ v < 30 m/s, SS304L
t ≧ 288 °C	SS316L	v < 30 m/s, SS304L or SS316L
		Atmospheric transfer line: SS304L or SS316L
		Vacuum transfer line: SS316L or SS317L
Remarks	1. t ≧ 240 °C and v ≧ 30 m/s: choosing SS316L with Mo% ≧ 2.5 %,or choosing SS317L, for example the atmospheric and vacuum transfer lines	
	2. When corrosion severely, material selection of atmospheric tower packing and vacuum tower packing supports can be improve a grade.	
	3. Material selection of vacuum tower packing improves a grade, considering packing is thinner and has more severe corrosion cases. When 240 °C ≤ t < 288 °C, choosing SS316L. When t ≧ 288 °C, choosing SS317L. When corrosion severely,SS317L could also be chosen when 240 °C ≤ t < 288 °C	
	4. SS304L can be replaced with SS304 or SS321. SS316L can be replaced with SS316;SS317L can be replaced with SS317	

2.2 Material Selecting of Chinese Refineries

For Chinese distillation units processing high-acid crude oil, equipment and piping materials of new or upgraded units basically meet the guideline SH/T 3129, for example Huizhou refinery of CNOOC [7, 8], Qingdao, Maoming, Guangzhou, Zhenhai and Jinling companies of Sinopec. Material of the old distillation units without upgrading is in lower material grade yet, which corroded severely, for example Liaohe company of PetroChina.

3 Corrosion in Several Refineries

3.1 A Company of Sinopec in East China (Company A)

Distillation unit in company A had been carried out a large scale material upgrading in 2009 and started operation in November, which material upgrading generally meets the guideline SH/T 3129. Company A mainly processes imported high-acid crude oil after the upgrading, such as Doba, Albacora, Kuito, Roncador, Dar Blend, Mrlim. The mean TAN of blending feedstock of distillation unit was about 2.0 mg KOH/g. Sulfur content of the feedstock had some changes, which mean

Table 2 Twice corrosion inspect in 2011 and 2012 of distillation unit

Position	Corrosion inspection in June 2011	Corrosion inspection in August 2012
Flash column	Made of carbon steel, operated at about 220–240 °C. Many carbon steel lining welds at feed inlet area were corroded penetration. Feed oil receiver thinned severely by NAC erosion-corrosion, the thinnest thickness was less than 1 mm	Replaced according to plan
Atmospheric column	Tower wall and trays of side cut 2 and below operating at high temperature were well	
Vacuum tower wall	Tower wall of side cut 2 and below operating at high temperature were well	Tower wall at side cut 3 extracting and returning site were severely corroded. Pits depth was about 1–1.5 mm. Tower wall was repaired using SS317L sheet. Feeding and bottom parts were in good condition
Vacuum tower packing	Tower packing at vacuum side cut 3 extracting and returning site was corroded severely, large area of which was scattered. SS316L packing at side cut 3 back to tower site was overall replaced with SS317L. Packing thickness increased from 0.2 mm to 0.25 mm. Corroded packing at side cut 3 was patched	Tower packing at vacuum side cut 3 was corroded severely, large area of which was scattered with 200–600 mm height and was overall replaced with material SS317L during this shutdown time. Packing at side cut 3 back to tower was basically well with only small localized corrosion
Vacuum tower internals	Packing support and liquid distributor were well. A shot piece of vacuum side cut 3 back to tower distributor was misused SS321, and occurred large area penetrated corrosion	Packing support, liquid distributor, floating ball level meter and bolts were severely pitted. The liquid distributor pipeline fillet welds of vacuum side cut 3 back to tower was found several penetrated corrosion. U-steel with SS316L for fixing liquid distributors was found several large area penetrated corrosion
Vacuum transfer line	Inner wall was smooth, and the weld joints were in good condition	Pipe weld joints and thermowell had obvious erosion-corrosion. The top half part inner wall had obvious pitting
Atmospheric and vacuum furnace tubes	The furnace radiation tubes thickness were measured three times in November 2010, June 2011 and August 2012. There were no obvious changes	
High temperature heat exchanges	Material grade was higher, no obvious corrosion was found	
High temperature piping with SS316L	There was no obvious corrosion was found	

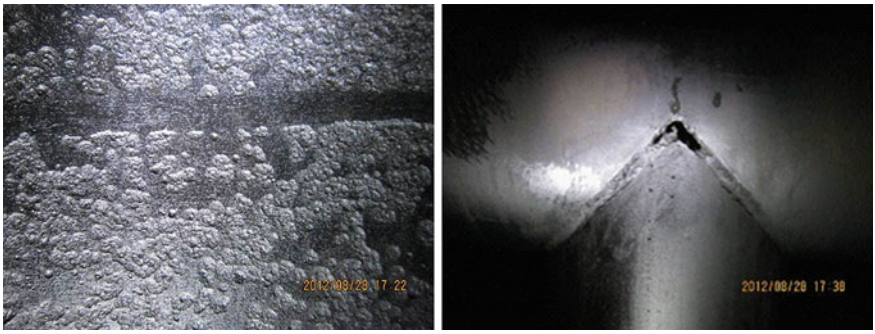


Fig. 1 Pitting corrosion of tower wall, fillet welds penetrated corrosion distributor pipeline at vacuum side cut 3 back to tower

value was 0.52 % before June 2011 and was up to 0.79 % from July 2011 to June 2012. Twice Corrosion inspection was conducted in the year 2011 for eliminating defects and 2012 for shutdown maintenance. Typical NAC problems were found, as shown in Table 2. Twice inspection both found that the worst parts of high temperature NAC were in the vacuum column where the vacuum side cut 3 extracting and returning site. The SS316L tower packing of vacuum side cut 3 back to tower was corroded severely and overall replaced with SS317L, and the thickness was increased to 0.25 from 0.2 mm in June 2011. The new tower packing was general well with only small localized corrosion in August 2012. The tower wall, oil collection tank, packing support and liquid distributor at vacuum side cut 3 extracting and returning site were generally well in 2011, but which were found corrosion severely in 2012. Tower wall at vacuum side cut 3 extracting and returning site were pitted densely, pits depth was about 1–1.5 mm, as shown in Fig. 1. Packing support, liquid distributor, floating ball level meter and bolts also had severe pitting. The liquid distributor pipeline fillet welds of vacuum side cut 3 back to tower had several penetrated corrosion, as shown in Fig. 1. U-steel with material SS316L for fixing liquid distributors were also found several large area penetrated corrosion. The pipe inner wall of vacuum transfer line was smooth, and the weld joints of which were well in 2011. However, it was found that the top half part inner wall of vacuum transfer line had obvious pitting and the pipe weld joints and thermowell had obvious erosion-corrosion in 2012.

3.2 A Company of PetroChina in Northeast China (Company B)

The distillation unit in company B was put into operation in 1993, mainly processing Chinese Liaohe crude oil. The unit mainly processed Liaohe blend crude oil and Venezuela Merey16 crude oil after January 2009. Crude oils properties were

Table 3 Crude oils properties [9]

Crude oil name	TAN mgKOH/g	Sulfur content (%)
Liaohe low solidifying point oil	3.85	0.14
Liaohe super heavy oil	5.53	0.51
Liaohe blend oil	1.72	0.11
Venezuela Merey16 crude oil	1.47	2.1

shown in Table 3. Liaohe crude oil was typically high-acid low-sulfur property, Venezuela Merey16 crude oil was typically high-acid high-sulfur property.

Corrosion inspection of distillation unit was conducted during shutdown maintenance time in August 2012. Material of high temperature parts of atmospheric and vacuum column was SS316L lining. Tower packing and trays were also SS316L. High temperature part of atmospheric column was well, no obvious corrosion found. Vacuum tower wall at side cut 3 to cut 5 had been found corrosion perforation, which were repaired with SS316L lining full circle in 2008 and was still well in 2012. Tower trays and packing operating at high temperature were well.

Atmospheric furnace convection and radiant tubes were 5Cr1/2Mo steel. Vacuum furnace convection tubes were 5Cr1/2Mo steel, and also the radiant inlet tubes. Vacuum furnace radiant outlet tubes were SS316L. They were all in good station after 3 years operation, no obvious corrosion thinning was found.

Some heat exchangers were upgraded to SS304 or SS316L in January 2009, but the left were still carbon steel. There was 35 heat exchangers operating temperature over 220 °C, 19 of which were found severe naphthenic acid corrosion. The following characteristics of NAC could be drawn from the inspection results.

(1) Carbon steel occurs significantly NAC as long as the temperature exceeding 220 °C, especially at the high flow rates and/or turbulent flow area. For example, typical erosion-corrosion occurred in E119A-D channels with the fluid of only 220–240 °C crude oil, as shown in Fig. 2.

Fig. 2 Erosion-corrosion at the junction of channel flange and partition plate of E119A (*left figure*), corrosion of tubes and tubesheet of E112A (*right figure*)

Table 4 Corrosion cases of PetroChina Jinzhou company

year	Corrosion cases
1995	Vacuum tower wall with SS321 lining leaked by corrosion
2001	Six vacuum residue—topped oil heat exchangers had internal leakage occurred one after another, causing plant shutdown
2002	Residue pump loop line of visbreaking unit was corroded cracking and fired
2004	Residual oil distributor pipe of visbreaking unit was corroded leaking and fired
2005	A heat exchanger shell nozzle of visbreaking unit leaked and fired
2006	Pump outlet pipeline of vacuum side cut 3 leaked and fired
2006	Outlet valve flange of vacuum side cut 4 was corroded perforation
2006	Thermocouples of vacuum furnace four feed all corrosion leaked
2006	Vacuum side cut 3 pump outlet 304SS pipeline thinned from 8 to 2 mm.
2007	A heat exchanger inlet nozzle of vacuum side cut 3 leaked and fired
2010	A flange neck of vacuum side cut 4 outlet pipeline thinned to 1 mm. Another one was material SS316L and no obvious corrosion

(2) Material SS304 used as heat exchanger tubes in atmospheric system was well, but which was found severe corrosion in vacuum system, especially when operating temperature exceeding 280 °C. For example tubes of E112A and E140, which channel fluid was vacuum side cut 3 and operating at about 290 °C, were found severe corrosion after used 3 years, as shown in Fig. 2. Tubes of E112B were corroded slightly, because the fluid temperature of which was lower after cooled by E112A. That also indicated the importance of temperature on the corrosion of SS304 tubes. So it's suggested that material SS316L should be used when NAC is severe.

PertroChina Jinxi company [10, 11] also processes high-acid low-sulfur Liaohe crude oil, corrosion leaks occurred many times as shown in Table 4. The distillation unit had been found corrosion leaks 140 times during the year 2004–2007, including corrosion in low temperature. Equipment nozzles and pipelines corrosion leaks by high temperature NAC amounted to dozens of times. The vacuum tower internals material was replaced to SS316L in 1990, after that the corrosion of SS321 vacuum tower wall increased. Tower wall of vacuum side cut 2 leaked by corrosion in 1995, which made a shutdown for repair. The vacuum tower wall was replaced by SS316L lining in 1996. The important equipment and pipelines were also material upgraded to SS316L during 2005 and 2007 shutdown time. Corrosion situation had significant improvement after that.

PertroChina Jinzhou company [12, 13] processes Liaohe crude oil too. Vacuum side cut 3 packing with materials Cr18-Ni8 was almost completely ineffective in less than a year in 1989. Almost all the local supports of vacuum side cut 2 and cut 3 was completely corroded. Vacuum side cut 3 packing was replaced with SS316L trays in 1990, which was bright as new after used 2 years. The vacuum tower wall was corrosion leaked 8 times in 1996–1999 and the most severe parts were side cut 2 and cut 3, which was replaced to SS316L lining. Vacuum furnace radiant tubes

for 5Cr1/2Mo steel leaked and fired in 2004. The leaked tubes were the first and second tubes just after diameter expansion. The vacuum furnace had four feedstock inlet ways, three of which leaked, another was also thinned.

3.3 A Company of Sinopec in South China (Company C)

Distillation unit processing capacity in company C was 3 million ton per year. It had been conducted a large scale material upgrading in 2007, which material selecting basically meeting SH/T 3129. The distillation unit mainly processed high-acid crude oil such as Dar Blend, Albacora. after the year 2007. The average TAN of processing crude oil was 1.62mgKOH/g and the average sulfur content was about 0.6 % in 2008.

The distillation unit was conducted shutdown maintenance in the end of 2008. The atmospheric column was well with no obvious corrosion. The vacuum column trays between side cut 2 and cut 3 were material SS321, back of which had obvious pitting, and were replaced by SS316L. The below four layer trays were SS316L and no obvious corrosion found.

Atmospheric furnace convection tubes were putted into using in 2007, which was material 5Cr1/2Mo and Φ152 × 8 mm size. The tubes were replaced to SS321 material according to plan during this maintenance time. The 5Cr1/2Mo furnace tubes straight parts were well. Thickness of most elbows were 7–7.3 mm, the corrosion rate calculated by thickness reduction was about 0.5 mm/year, closed to the estimated value according to API 581.

Vacuum residue pipeline from tower outlet to pump inlet was 5Cr1/2Mo steel and operating temperature was 350 °C, which was corrosion thinning widely. The thickness difference of the thinnest and thickest area was up to about 3 mm. Pipeline inner surface had density pits. The pipeline was replaced to material SS316L.

Atmospheric high-speed transfer line with SS321 material had been found a corrosion rate up to about 0.3 mm/year by fixed-point thickness measurement on line, which was replaced to material SS316L during maintenance time in the end of 2008.

Shell nozzles of two heat exchangers operating at 220–240 °C with fluid of topped oil and residue oil had been found obvious corrosion thinning, which was typical NAC for carbon steel.

4 Corrosion Monitoring

High temperature naphthenic acid corrosion could easily lead to corrosion thinning or even localized corrosion perforation. Once leaking would result in fire, explosion and other major security incidents. Therefore, corrosion monitoring during plant operation is significant. Chinese companies processing high-acid crude oil now can

make use of fixed-point thickness measurement, online probe, hydrogen flux measurement, distillate analysis including TAN, sulfur content, Fe and Ni ion, which help to comprehensively analyze the unit corrosion condition. New and upgraded units had more comprehensive monitoring measurements. Some old units were more a lack of which.

China Sinopec Qingdao company distillation unit conducted material upgrading in 2009. 17 online monitoring probes were installed, 9 of which used to monitor NAC. CNOOC Huizhou refinery considered the online monitoring program in the design stage. 17 online probes were installed before the distillation unit going into production in 2009, 10 of which used to monitor NAC [8]. China Sinopec Guangzhou company distillation unit not only used portable hydrogen flux measurement, an online hydrogen flux probe was also installed at the vacuum side cut 3 outlet pipeline.

In 2007, four atmospheric and vacuum distillation units in Jinling, Guangzhou, Zhenhai, Maoming company of Sinopec conducted material upgrading for adapting to process high-acid crude oil such as Sudan Dar Bland et al. Hefei General Machinery Research Institute (HGMRI) conducted corrosion monitoring for 2 years and a half from 2008 to 2010. Materials corrosion data for NAC were accumulated. HGMRI also conducted a long term corrosion monitoring for Sinopec Qingdao company.

There was no significant NAC for carbon steel pipeline operating temperature below 220 °C, corrosion rate of which was less than 0.25 mm/year. Some pipeline operating temperature between 220 ~ 240 °C with high TAN fluid occurred obvious corrosion. Some carbon steel pipeline had a corrosion rate of about 0.3 mm/year, which was corresponding to the corrosion of heat exchanger channel in company B and heat exchanger shell nozzle thinning in company C.

Corrosion of most SS321 and SS304 piping was not obvious when operating temperature below 288 °C, which corrosion rate was ≤0.1 mm/year. However, piping exhibited obvious corrosion thinning trend at higher temperature and higher flow rate. Material SS321 atmospheric transfer line had a corrosion rate of 0.3 mm/year, which fluid temperature was 360 °C with a flow rate of about 20–30 m/s.

High temperature tower side cuts with material SS316L had no obvious corrosion thinning, the corrosion rate of most of which was ≤0.1 mm/year. A few of high temperature and high velocity transfer lines had a corrosion rate of about 0.1–0.2 mm/year, which was corresponding to the corrosion of vacuum transfer line found in company A.

5 Analysis and Discussion

5.1 Corrosion Analysis of Vacuum Column

Corrosion resistance of material SS316L used in the general equipment and piping is better, but which encountered severe corrosion problems when used as vacuum tower lining, structured packing and other internals, even SS317L structured

packing also corrosion severe. For example, material SS316L vacuum tower lining and packing in company A, SS317L structured packing in another Sinopec company had found severe corrosion. Vacuum column corrosion was closely related to the crude oil types, TAN and sulfur content. Company A had amazing corrosion after 3 years running. Huizhou refinery of CNOOC processed higher TAN Chinese Penglai crude oil, which TAN was 3.46 mg KOH/g, sulfur content was 0.29 %. It was found well after running an operation period, only tower packing of vacuum side cut 3 had little localized corrosion [14]. Company B, Jinzhou and Jinxi companies of PetroChina all processed high-acid low-sulfur Liaohe crude oil. Processing experiences showed that material 316L were applied in good condition.

Gutzeit reported the results of field surveys which indicate that naphthenic acid corrosion is greatest at the condensation point where the vapour stream draws liquid over the metal. Scattergood et al. have reported that the observed corrosion is most severe at the liquid/vapour interface where the vapour forms a liquid film over the metal surface. Blanco noted that the severity of corrosion appeared to be higher when the physical state of the acids was changing, for instance in a vaporisation situation such as a transfer line or in a condensing situation such as the vacuum column [2]. In vacuum processing environment, the most severe corrosion often occurs at approximately 288 °C, i.e. the site of vacuum side cut 3. The reason is that the naphthenic acids tend to concentrate in those fluids with true boiling points in the 370–425 °C range, and the effect of a vacuum (in the vacuum column) is to reduce the boiling point 110–160 °C [15].

The vacuum tower lining and internals made of SS316L in company A was well when running a year and a half, but severe corrosion occurred after running about 3 years. There may be two reasons responsible for the corrosion. One was the influence of crude oils sulfur content which increased from 0.5 to 0.79 %, which promoted the corrosion. Another possible reason may be that the corrosion of material SS316L has an incubation stage. Corrosion is slower in the initial stage, following by a rapid development once the corrosion forming which made the surface of the metal rough. Therefore, every time a unit shutdown, corrosion inspection should be carefully conducted.

5.2 Corrosion Analysis of Furnace Tubes and Transfer Lines

Atmospheric furnace tubes, vacuum furnace convection tubes and radiation tubes just before diameter expansion had fewer corrosion cases. Even furnace tubes with lower grade material were generally used in good condition. For example, the furnace tubes in above position of company B were material 5Cr1/2Mo steel, which had no obvious thinning after running a period.

However, vacuum furnace radiation tubes after diameter expansion and vacuum transfer lines had obvious corrosion. For example, Karamay petrochemical company processes high-acid crude oil. The vacuum furnace tubes with carbon steel before diameter expansion had a corrosion rate of 0.5 mm/year, but the tubes after

diameter expansion had a corrosion rate up to 10 mm/year. In 2004, Jinzhou petrochemical 5Cr1/2Mo steel vacuum furnace radiation tubes leaked by corrosion, the leak was located just after the tubes diameter expansion [13]. For another example, SS316L vacuum transfer line in company A had obvious pitting and erosion-corrosion.

High temperature naphthenic acid corrosion is influenced by naphthenic acid molecule activation kinetics, naphthenic acid molecules adsorption on the metal surface. Fluid in furnace tubes is constantly heat absorbing and constantly gasifying, which makes the process of naphthenic acid molecules adsorbed on the metal surface inhibited, and then the tubes corrosion is not sever. However, vacuum furnace radiation tubes after diameter expansion and vacuum transfer lines are in the state of fluid gasification sudden increasing and flow rate mutation, which makes the high temperature sulfur corrosion and naphthenic acid erosion-corrosion greatly increased.

5.3 Others

Carbon steel is not applied when operating temperature over 220 °C. The corrosion monitoring, corrosion cases of flash column in company A and heat exchanger shell thinning in company C all illustrate this view.

SS321 and SS304 stainless steel may not be applied to equipment operating temperature over 280 °C in some cases. For example, SS304 heat exchanger tubes had obvious corrosion in company B. They are not applied to pipelines for high temperature and high fluid velocity too. For example, atmospheric transfer line had obvious corrosion thinning in company C.

The units processing low-acid crude oil should be timely assessed and conduct corrosion monitoring, even blending a small amount of acid crude oil. It should be material upgraded at an appropriate time. For example, a company in North China processed low-sulfur and low-acid crude oil. The TAN increased to 0.3–0.4 mg KOH/g when blending some Jidong crude oil later. The residue pump outlet pipeline had a big fire caused by corrosion leaking, which made the unit a shutdown for 16 days in 2009 [1].

6 Summary and Recommendation

Corrosion and protection in China refineries processing high-acid crude have made significant progress. High temperature naphthenic acid corrosion is generally able to be control through the reasonable material selection. Corrosion of vacuum tower wall, packing and other internals is the main bottleneck for units' long-term operation. Integrated corrosion protection can be conducted by choosing SS317L,

increasing packing thickness, and the injection of corrosion inhibitor if necessary judging from corrosion monitoring.

For the units processing crude oil that TAN rising above the design value and haven't conducted material upgrading, it was suggested to do corrosion evaluation and corrosion monitoring. The corrosion sensitive parts should be conducted material upgrading during the units shutdown time.

It is recommended to conduct corrosion inspection during each shutdown time, especially atmospheric and vacuum columns, transfer lines, furnace tubes, and equipment and piping with lower grade materials. Corrosion status changes could be mastered through corrosion inspection. Material upgrading should be conducted when necessary.

References

1. Ren G (2010) Corrosion and protection of oil refineries processing high-acid crude oil. Petrochem Equipment Technol 31(4):54–57
2. Slavcheva E, Shone B, Turnbull A (1999) Review of naphthenic acid corrosion in oil refining. Br Corros J 34(2):125–131
3. Tebba S, Kane RD (1998) Assessment of crude oil corrosivity. In: Corrosion 98, NACE international conference, Houston, Paper 578
4. Kane RD, Cayard MS (2002) A comprehensive study on naphthenic acid corrosion. In: Corrosion 2002, NACE international conference, Houston, Paper 02555
5. Groysman A, Brodsky B, Pener J et al (2007) Low temperature naphthenic acid corrosion study. In: Corrosion 2007, NACE international conference, Houston. Paper 07569
6. Dettman HD, Li N, Luo J (2009) Refinery corrosion, organic acid structure, and athabasca bitumen. In: Corrosion 2009, NACE international conference, Houston, Paper 09336
7. Xia CP (2010) Corrosion protection technical features of distillation unit processing high acid heavy crude oil. Petrochem Equipment Technol 31(6):57–61
8. Ouyang J (2011) Design for corrosion protection in crude distillation unit processing high-acidity crude. Corros Prot Petrochem Ind 28(2):35–39
9. Zhao XK (2011) Equipment corrosion and protection on processing high-acid crude oil. Chem Eng Equipment 6:90–91
10. Zhao Y, An H (2002) Equipment corrosion in crude distillation and visbreaking complex unit and protection. Corros Prot Petrochem Ind 19(4):12–17
11. Jiang YQ (2011) Experience and discussion on corrosion protection of distillation unit. Petrochem Equipment Technol 32(1):22–24
12. Si ZP, Chen WY (2000) Application of SS316L in vacuum column. Corros Prot Petrochem Ind 17(4):29–33
13. Xu L (2008) Study on sour crude oil processing. Corros Prot Petrochem Ind 25(6):21–25
14. Sun L, Zheng GM, Zhang JF (2012) Study on naphthenic acid corrosion of crude oil from Penglai 19-3 well. Corros Prot Petrochem Ind 29(6):8–10
15. White RA (1998) Materials selection for petroleum refineries and gathering facilities. NACE International, Houston, pp 17

FEM Simulation of Nonlinear Lamb Waves for Detecting a Micro-Crack in a Metallic Plate

Xiang Wan, Peter W. Tse, Guanghua Xu, Tangfei Tao, Fei Liu, Xiaoguang Chen and Qing Zhang

Abstract Nonlinear ultrasonic technique has been employed to detect micro-cracks since conventional linear elastic ultrasonic technology is just sensitive to gross defects. However, most of nonlinear ultrasonic researches to date have been experimental generally using bulk waves or Rayleigh waves, few numerical studies exist, especially for lamb wave ultrasonic. In this chapter, finite element method (FEM) is applied to simulate nonlinear lamb waves interacting with a micro-crack in a thin metallic structure. A pitch and catch approach is introduced containing two symmetric piezoelectric transition (PZT) wafers as actuators to generate single S0 mode signal and one PZT wafer as receiver. Generated S0 mode lamb waves propagate along the structure, interact with the micro-crack, obtain nonlinear features, and are picked up by the receiver. An undamaged plate and seven plates with different crack length are simulated. The received simulation signal from a micro-cracked plate contains a S0 mode wave-packet and a new wave-packet. A nonlinearity index (NI) is proposed to show the degree of nonlinear effect. The simulation results show that employing NI, a received signal of S0 mode can provide information on the damage severity of a micro-crack and the new wave-packet signal can be used as an early indicator for the existence of a micro-crack.

Keywords FEM simulation · Nonlinear lamb waves · Micro-cracks

X. Wan · G. Xu (✉) · T. Tao · F. Liu · X. Chen · Q. Zhang
School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China
e-mail: xugh@mail.xjtu.edu.cn

G. Xu
State Key Laboratory for Manufacturing Systems Engineering,
Xi'an Jiaotong University, Xi'an 710049, China

P.W. Tse
Department of Systems Engineering and Engineering Management,
City University of Hong Kong, Tat Chee Avenue,
Kowloon, Hong Kong, China

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_133

1 Introduction

Inspection and evaluation of structural changes of materials to ensure their structural integrity is obviously essential in various industries. Especially detection of incipient micro-cracks at the early stage of fracture is of increasing importance in vital components in aircrafts, unclear power plants, chemical plants and refinery plants in order to ensure their safety.

As one of the most powerful non-destructive evaluation (NDE) tools, conventional ultrasonic testing method has been extensively utilized to effectively detect and size a defect with volume, like an open crack or a void. However, traditional ultrasonic technology based on linear theory is just sensitive to gross damages and is unable to detect micro-cracks. An alternative way to overcome this limitation is employing nonlinear ultrasonic technology. The nonlinear ultrasonic technology, which uses distinctive harmonics features, proves itself a promising approach to detect micro-cracks [1]. Contact acoustic nonlinearity (CAN) is a kind of physical mechanism of higher harmonic generation and it is suitable for detection of micro-cracks.

Nonlinear guided wave technology is of great interest as they combine the high sensitivity of nonlinear approach with large testing ranges of guided waves. To date, most of researches have been experimental generally using bulk waves or Rayleigh waves [2, 3]. Experimental investigation for evaluating fatigue micro-cracks using nonlinear lamb waves has been conducted recently [4]. Only a few investigations using finite element method exist. Kawashima et al. [5] studied the CAN utilizing Rayleigh waves to detect surface cracks. Soshu and Toshihiko [6] employed nonlinear longitudinal waves to detect a closed crack. Recently, Yanfeng Shen and Giurgiutiu [7] modelled a piezoelectric wafer active sensor (PWAS) to excite both S0 and A0 mode to interrogate a plate with a breathing-crack on the surface. However, breathing cracks modeling can not accurately represent real cracks in a component in unclear power plants, chemical plants or refinery plants in most cases.

In this study, FEM is employed to simulate nonlinear lamb waves interacting with a closed micro-crack in a metallic plate. The theory of CAN is briefly explained as a foundation of detecting micro-cracks using nonlinear ultrasonics. In the FEM model, a pitch and catch approach is employed containing two symmetric PZT wafers as transmitters generate single S0 mode signal and a single PZT wafer as receiver, and the modeling of a closed micro-crack with oval shape is simulated by hard contact with frictionless model. Next, FEM simulation results are displayed and discussed. Conclusions are drawn and future studies are proposed at last.

2 Generation of Higher Harmonics Through CAN

When ultrasonic wave excited by large amplitude is incident to imperfect interface, higher harmonic waves are generated. This kind of phenomenon is so-called as contact acoustic nonlinearity (CAN), and it is recently attracting increasing attention regarding the characterization of closed cracks or imperfect bond interfaces. This effect originates in repetition of collisions between the two surface caused by incident ultrasonic waves. When ultrasonic waves reach an imperfect contact interface, the compressional part of the waves can penetrate it, but their tensile part can not penetrate it, as shown in Fig. 1 [8]. Therefore, after penetrating the interface, the waves become nearly half-wave rectification, which means they have obvious nonlinearity, and this nonlinearity can be detected by higher harmonics [1].

3 Finite Element Method Model

Finite element model is shown in Fig. 2. In our model, we consider one half of the plate symmetric about the y-axis. The plate is long enough to ensure the received waves are not affected by the plate boundary reflections. Three 6.4 mm × 6.4 mm × 0.1 mm PZT wafers are bonded on a 2 mm thick aluminum plate. Two PZT wafers used as actuators are placed on the double surfaces of the exciting point which indicates these two PZT transmitters have the same coordinate. When these two PZT actuators are excited by the same input burst signal, symmetric mode lamb waves should be enhanced while anti-symmetric ones should be weakened. In our case, just S0 mode lamb waves are generated. Another one PZT wafer function as receiver is placed to collect wave signal.

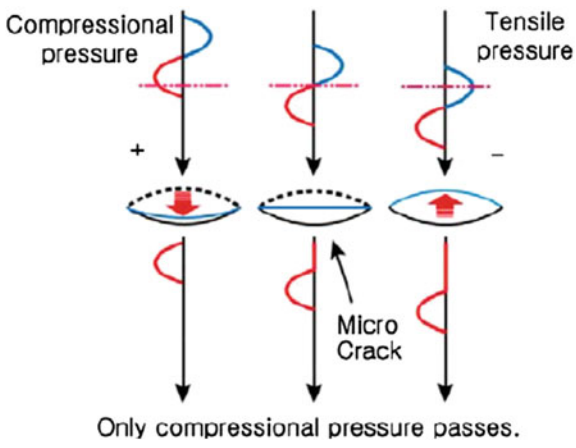


Fig. 1 The schematic diagram showing the concept of CAN at micro-crack [8]

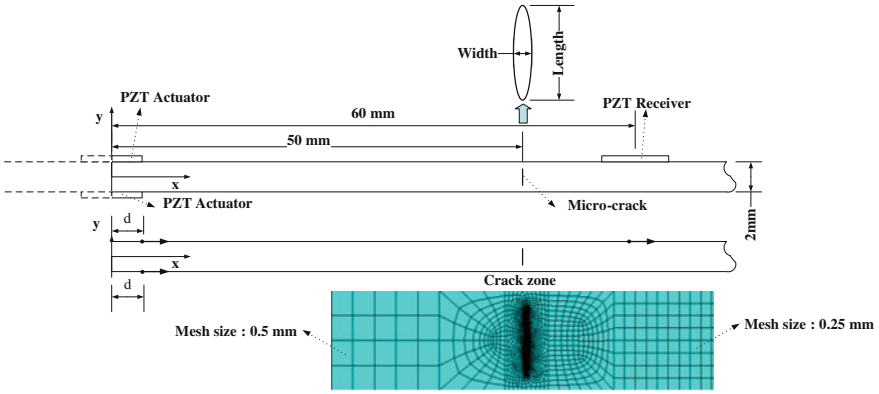


Fig. 2 The finite element model

We employ point source to model PZT actuators as shown in Fig. 2. This kind of source point modeling is described in detail in [9]. Two point sources are applied to the upper surface and the lower surface respectively a distance d from the y axis where d equals to a half length of a PZT transducer. A hamming windowed tone-burst consisting of five cycles at the frequency of 400 kHz is utilized as excitation signal and the time domain waveform and its fast fourier transform (FFT) spectrum are depicted in Fig. 3.

The micro-crack is located at 50 mm from the y axis. The shape of the micro-crack is modeled as ellipse shown in Fig. 2. In our simulation, the micro-crack is considered with a width of 6 nm and a length of 200, 400, 600, 800, 1000, 1200 and 1400 um. Here, we use an index s = l/t (where l and t refer to the crack length and plate thickness respectively) to define the micro-crack severity. Accordingly, the micro-crack severity index s equals to 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 and 0.7. For

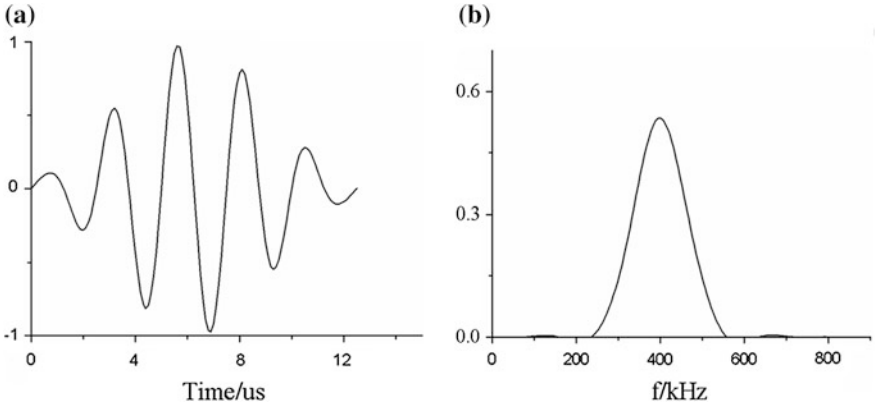


Fig. 3 The excitation signal in **a** time domain and **b** frequency domain

undamaged plate, the micro-crack severity index s is 0. We assumed aluminum plate, E = 69 GPa, ν = 0.33 and ρ = 2700 kg/m^3. The PZT receiver is modeled by a point on the upper surface located at 60 mm from the y axis. And the stress σ_{11} at this point is monitored as the receiving signal.

Abaqus Explicit method is used in this study. In order to obtain adequate accuracy and high efficiency, a meshing strategy with varying mesh density is performed. In general, a denser mesh will give a more accurate result, but will also cost more calculation time and computer resources. The maximum element size and time step to ensure accuracy is adopted in [10]:

$$I_{\max} = \frac{\lambda_{\min}}{20} \tag{1}$$

$$\Delta t_{\max} = \frac{1}{20 f_{\max}} \tag{2}$$

For 400 kHz signal, a mesh size of 0.5 mm and a time step of 0.1 us are enough to ensure accuracy according to Eqs. (1) and (2). Therefore, a mesh size of 0.5 mm is applied between the actuators and the micro-crack. The crack zone is more densely meshed with much smaller elements because of complicated mechanical response at the crack zone. After interacting with the micro-crack, higher frequency components will be generated. In order to ensure its accuracy, a mesh size of 0.25 mm is used between the crack zone and the receiver. And the meshing result is depicted in Fig. 2. The time step of 0.05 us is adopted to ensure the accuracy of higher harmonics.

4 Simulation Results and Discussion

The received time domain signal is shown in Fig. 4a for undamaged case (s = 0) and in Fig. 4b for a cracked case (s = 0.5). The superposed time domain simulation signals for both undamaged case (s = 0) and the micro-cracked case (s = 0.5) are depicted in Fig. 5. It can be obviously noted that there is a slight amplitude drop and phase shift of received signal from the micro-cracked plate. Another big difference is the presence of a new wave-packet as a result of the existence of the micro-crack in the plate. This new wave packet may be introduced by mode conversion [7] and the contact nonlinearity effect at the micro-crack. Higher harmonics components will also be introduced into S0 wave-packet, as it is dispersive. It can be deduced that the new wave-packet originates from the micro-crack and mode conversion from S0 mode. Fourier transform of S0 mode and the new wave-packet signal for the cracked plate are conducted and plotted in Fig. 6.

For undamaged plate, no higher harmonics components exist in either S0 mode or the new wave-packet whereas nonlinear second harmonic are obviously shown in the signals from the micro-cracked plate. FFT spectrum of S0 mode is shown in

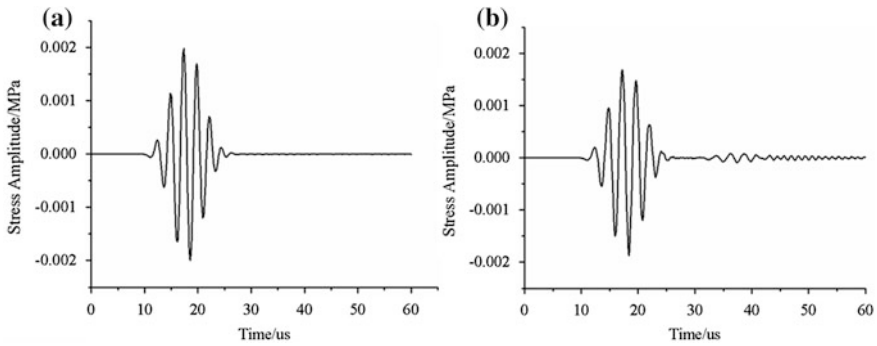


Fig. 4 The time domain simulation signal for **a** undamaged case (s = 0) and **b** cracked case (s = 0.5)

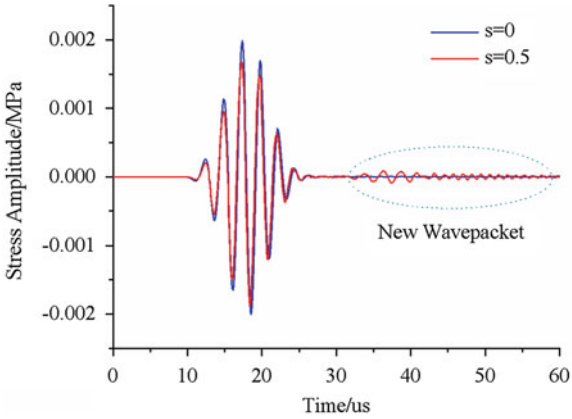


Fig. 5 Superposed time domain simulation signals for undamaged case (s = 0) and cracked case (s = 0.5)

Fig. 6a. Since the fundamental excitation frequency is centred at f = 400 kHz, the 398.7 kHz peak corresponds to the fundamental frequency f of excitation signal, and the 830.6 kHz correspond to the second harmonic 2f. For FFT spectrum of the new wave-packet signal depicted in Fig. 6b, the first peak at 433.3 kHz corresponds to the fundamental frequency f and the second harmonic component 2f is clearly observed at 833.3 kHz. In the new wave-packet, the amplitude of nonlinear second harmonic is more obvious and is closer to that of the excitation than S0 mode signal.

The amplitudes of second harmonic 2f for both S0 mode and the new wave-packet with different micro-crack severities r = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 and 0.7 are shown in Fig. 7. It can be obviously observed that though the amplitude of the second harmonic 2f is very small, it has a monotonically increasing relationship

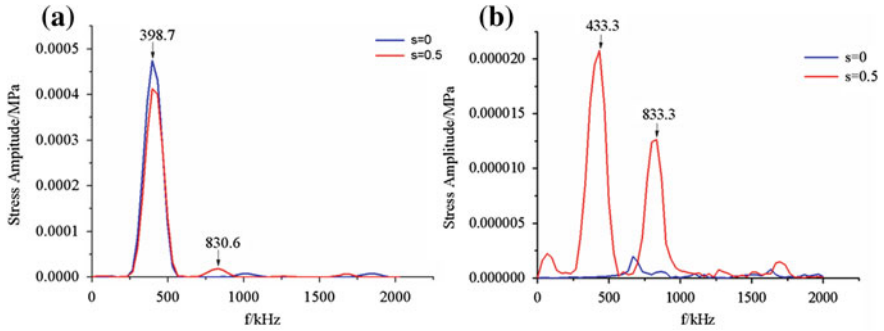


Fig. 6 The Fourier transform of S0 and the new wave-packet for **a** S0 mode and **b** new wave-packet

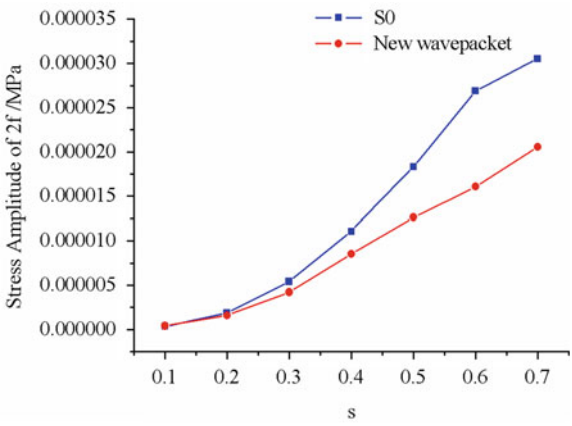


Fig. 7 The amplitudes of 2f for S0 and new wave-packet with the micro-crack severity index

with the micro-crack severity index for both S0 wave and the new wave-packet. The presence of the micro-crack introducing CAN to both S0 wave signal and the new wave-packet is clearly presented.

We employed the amplitude ratio of second harmonic signal to fundamental frequency signal detected by the PZT receiver as a nonlinear index (NI) to present the degree of nonlinear effect and it is stated in Eq. (3) where Amp (f) and Amp (2f) refer to the FFT spectrum amplitude at the fundamental frequency and at the second harmonic frequency.

$$NI = \frac{Amp(2f)}{Amp(f)} \tag{3}$$

The trend of NI with the micro-crack severity index is shown for S0 mode in Fig. 8a and for the new wave-packet in Fig. 8b. It is clearly noted that the NI is relatively small for S0 mode signal compared with the new wave-packet. The NI for S0 mode is increased monotonically with the micro-crack severity index. However,

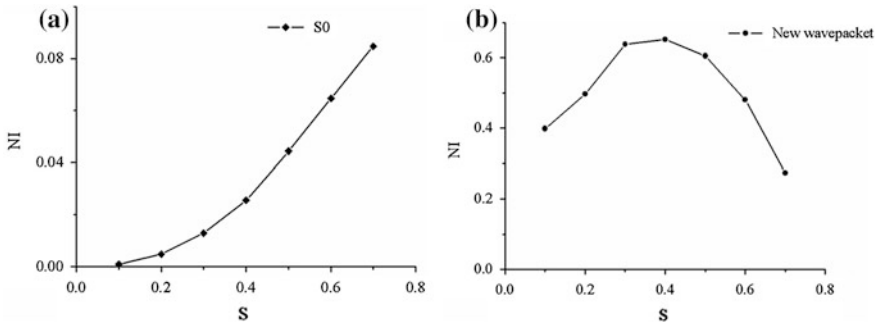


Fig. 8 Nonlinearity index (NI) with micro-crack severity index s for **a** S0 mode and **b** new wave-packet

the NI for the new wave-packet increases monotonically to a peak value and then decreases monotonically. Therefore, the NI for S0 mode wave signal can function as an indicator of damage severity of a micro-crack, while the NI for the new wave-packet can server as an early indicator for the existence of a micro-crack.

5 Conclusions

In this chapter, FEM simulation method is proposed to study nonlinear S0 mode lamb waves interacting with a micro-crack in a metallic plate. The simulation result shows that the presence of a new wave-packet indicates the existence of a micro-crack. From FFT spectrum analysis, no higher harmonics exist in either S0 mode or new wave-packet for undamaged plate whereas nonlinear second harmonic can be clearly observed in both S0 mode and the new wave-packet signals from the micro-cracked plate. And the amplitude of nonlinear second harmonics shows a monotonous relationship with the micro-crack severity index. In order to present the degree of nonlinear effect, a nonlinear index is adopted. The NI for S0 mode signal is found to increase monotonically with the micro-crack severity index. However, the NI for the new wave-packet is just very sensitive to low micro-crack severity index values (short length of micro-cracks). Employing NI, S0 mode signal can provide efficient information on the damage severity of a micro-crack and the new wave-packet signal can be used as an early indicator for the existence of a micro-crack.

The length of a micro-crack is just considered and discussed in this study. Other parameters, e.g. the width of a micro-crack, the shape of a micro-crack and the number of micro-cracks will be studied in the future work. Experimental investigations will also be performed to verify results from FEM simulation.

Acknowledgement This article was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 122011), a grant from City University of Hong Kong (Project No. 7008187) and a project of Xi'an Jiaotong University (Project No. 2012ZX04001-012-04).

References

1. Jhang KY (2009) Nonlinear ultrasonic techniques for non-linear destructive assessment of micro damage in material: a review. Int J Precis Eng Manuf 10(1):123–135
2. Morris WL, Buck O, Inman RV (1979) Acoustic harmonic generation due to fatigue damage in high-strength aluminum. J Appl Phys 50(11):6737–6741
3. Ogi H, Hirao M, Aoki S (2001) Non-contact monitoring of surface-wave nonlinearity for predicting the remaining life of fatigued steels. J Appl Phys 90(1):438–442
4. Dutta D, Sohn H, Harries KA (2009) A nonlinear acoustic technique for crack detection in metallic structures. Struct Health Monit Int J 8(6):573
5. Kawashima K, Omote R, Ito T, Fujita H, Shima T (2002) Nonlinear acoustic response through minute surface cracks: FEM simulation and experimentation. Ultrasonics 40(1):611–615
6. Soshu H, Toshihiko S (2006) Detection of a closed crack by nonlinear acoustics using ultrasonic transducers. In: Thompson DO, Chimenti, DE (eds) Review of quantitative non-destructive evaluation, pp 277–282. New York
7. Shen Y, Giurgiutiu V (2012) Predictive simulation of nonlinear ultrasonics. In: SPIE smart structures and materials + nondestructive evalucation and health monitoring. San Diego, California
8. Sutin A (1996) Nonlinear acoustic nondestructive testing of cracks.In: Proceedings of 14th International Symposium on Nonlinear Acoustics, pp 328–333)
9. Nieuwenhuis J, Neumann J, Greve D, Oppenheim I (2005) Generation and detection of guided waves using PZT wafer transducers. IEEE Trans Ultrason Ferroelectr Freq Control 52 (11):2103–2111
10. Giurgiutiu V (2005) Tuned Lamb wave excitation and detection with piezoelectric wafer active sensors for structural health monitoring. J Intell Mater Syst Struct 16(4):291–305

Introduction of the Risk Based Optimization and Risk Criteria Analysis of Spare Inventory in Petrochemical Plant

Jianxin Zhu, Wenbin Yuan, Peng Xu, Yunrong Lu and Xuedong Chen

Abstract Spares plays an important role in the asset integrity management. In petrochemical industry, great amount of money is occupied by spares which are lower inventory but higher unit price (critical spares). Some typical inventory optimization models, such as Economic Order Quantity (EOQ) method, which focuses mainly on the optimization of storage and order costs, is unsuited for the optimization of inventory of critical spares. While consume based analysis method can precisely predict the demand of spares only if the historical consuming data is adequate and sufficient. But in petrochemical industry, it is always difficult to obtain such data. In order to quantify the optimal spare inventory and determine stock strategy of critical spares, a risk based inventory analysis methodology was put forwarded in this chapter. The probability of stockout was obtained by systematic consideration of spare parts reliability, configuration of available spare equipments and spare parts and lead time of order. Taking stockout-costed loss into consideration, the risks of all critical spares of pumps in Y Company were obtained. It was found that the quantity of low risk spares whose potential yearly loss are less than 500,000 RMB take up to 80 %, while it account for 83 % of overall stock fund. It is except that great amount of money can be saved if those spares can be optimized by risk based spare optimization method. By comparing of the amount of spares and associated proportion of fund its' occupied, the determination of the risk criteria in spare inventory optimization was also discussed in the last part of this chapter.

Keyword Risk · Spare inventory · Optimization · Petrochemical installation · Risk criteria

J. Zhu (✉) · W. Yuan · P. Xu · Y. Lu · X. Chen
National Engineering and Technology Research Center on Pressure Vessel and Piping Safety, Hefei General Machinery Research Institute, No. 888, West Changjiang Rd, Hefei, China
e-mail: jxzhuhz@hotmail.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_134

1 Introduction

The spare parts inventory is of vital importance for the safety and availability of process installations due to the fact that diverse equipments from up and downstream sectors are closely interrelated. Great amount of money is spent on spare parts inventory in petrochemical installations due to the quantity and categories of diverse equipments involved. Statistics shows that the amount of money spent for spare inventory in a medium sized petrochemical enterprise (almost 50 installations) is up to 400–600 millions RMB, which account for 4–5 % of the Replacement Asset (RA). The number is too bigger compared with that of the world leading petrochemical enterprise, where spares account for only 2–2.8 % of RA. The oversized spare inventory in petrochemical enterprises brings not only system availability, but also other negative side effects: (1) Those badly required spares cannot be promptly updated because too much money was occupied in the oversized stock, which in turn leading to the unavailability of several important spare parts; (2) The amount of money spent in oversized stock may cause unnecessary interest loss. (3) Large quantity of spares has to be abandoned due to the wear out of spares in stock, which causes severe economic loss.

Nowadays more and more efforts were spent on the optimization of spare inventory. Diverse methodologies were applied in stock optimization. Generally it is believed that the Economic Ordering Quantity (EOQ) model is acceptable for the optimization of most easily worn parts (Class C and part of Class B spare parts) [1]. It is a pity, however that the efforts of such optimization is not always cost effective since the fund occupied by such spares is only take up to 20 percent, though quantitatively they account for 80 % of spares [2]. Since most Class C spares are low unit price, it seems that the management of Class A spares and part of Class B spares is the most cost effective as they account for nearly 80 % of total fund occupied, though numerically they only account for 10–20 % in quantity. The main considerations in EOQ model are purchase cost and storage cost. For those slow moving spares, the carrying charge is the most important factor to be considered. So the main consideration of spare optimization remain is, whether is it needed to store expensive spares and how to do?

The problem in storage strategy for slow moving spares is the evaluation of the consequence of stockout and associated probability (or risk). So ultimately the decision of storage of spare parts of category A and part of B rely mainly on the risk. In recent decades, the development of engineering risk management methodology in petrochemical industry has great improved. Own to its merit of coordinating between safety and economical efficiency, engineering risk evaluation methodology is widely accepted as a powerful tool in asset integrity management in petrochemical sector. Detailed engineering risk assessment method, such as Risk Based Inspection (RBI), Reliability Centered Maintenance (RCM) and Safety Integrity Level (SIL), are widely applied both in aboard and domestic petrochemical industry [3].

In this chapter, a novel risk based method was introduced in the optimization of inventory of slow moving spare parts. The risks of spare parts used in important pumps were evaluated quantitatively and the associated risk acceptance criterion was analyzed and discussed.

2 Propose of Risk Based Spare Inventory Optimization Method

Literary risk is a combination of both probability and consequence. Since diverse spare parts may have quite difference characteristics, in risk based optimization method, two main problems to be solved are which spares should be quantitatively evaluated and how to do?

In order to answer the first question, the spare inventory of Y Company was analyzed. Four categories of spares, namely bearing spares, specialized petrochemical spares, hydraulic equipment spares and sealing spares were quantitatively analyzed. The number of spares involved of all these 4 categories are 1449, 4061, 9896 and 11878, respectively. According to the enterprise regulations, spares belong to class A–C are listed in Table 1.

It is shown from Table 1 that spares belong to Class C take up to 75.02, 69.27, 81.27 and 90.31 %, respectively, which means most spare own the characteristics of low unit price and fast moving. Unlike Class C spares, quantitatively, spare parts belong to Class A only accounted for less than 15 %.

Investigation of stock fund involved in spare inventory management shown that those spares belong to Class A account for nearly 70–80 % of total fund occupation, though numerically they only take up to very few quantity. Such type of spares has the characteristics of low consumption, high unit price and unstable demand rate. Table 2 shows the distribution of the proportion of the quantity and fund occupied by two types of low consumption spares.

It is shown from Table 2 that most spares belongs to low demand rate spares (have a demand over 3 years). For such type of spares, those demand predict algorithm, such as moving average (MV), Exponential Smoothing (ES), Croston and Bootstrap sample method, are improper in demand prediction [4–6]. Due to lack of adequate consumption data, the optimization of such slow moving or nonmoving spares is time consuming [7].

Table 1 The quantity proportion of diverse spares in Y Company

Spare category	Class A (%)	Class B (%)	Class C (%)
Bearing spare category	4.21	20.77	75.02
Specialized spares category	11.70	19.03	69.27
Hydraulic spare category	4.85	13.88	81.27
Sealing spares category	1.18	8.51	90.31

Table 2 Distribution of the quantity and fund occupation of low consumption spares

Demand rate (Year)	Number of spares	Fund occupied (RMB)	Occupancy ratio (Quantity) (%)	Occupancy ratio (Fund) (%)
1–2 years	1722	40827998.6	6.18	9.32
Over 3 years	19422	226332538.1	69.72	51.64

In order to solve this problem, based on experience obtained by massive engineering risk assessment practices, a risk based spare inventory optimization method was proposed in this paper, risk accept criteria was also analyzed and discussed to evaluate the necessity of inventory.

3 Introduction of the Risk Based Spare Optimization Method

Risk management is a discipline that study both risk development and risk control [8]. In risk based spare optimization method, several procedures are involved in the overall risk evaluation methodology, namely analysis of stock influences, study of the probability of stockout, risk assessment, systematic risk controlling et al. (See Fig. 1 for details).

Detailed risk based spare optimization procedure is as follows: (1) Determining the classification of spares by the characteristics, criticality, unit price and consuming feature of diverse spares. Only those slow moving spares in Class A and

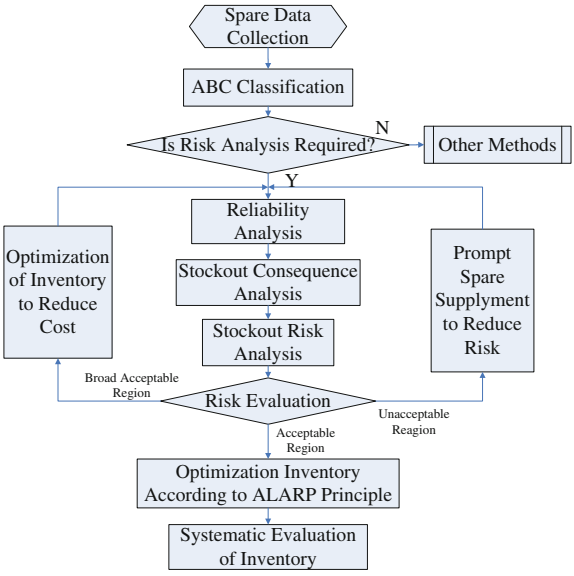


Fig. 1 Flowchart of risk based spare optimization method

part of Class B are to be optimized by this method; (2) Evaluating the probability and consequence of stockout by taking into consideration of economic loss caused by downtime and production interruption. The probability of out of stock should consider unit failure probability, spare equipment and spare parts configuration. The share of common spare by several machines should also be considered in determining failure probability. (3) Determination of spare optimization strategy. Risk acceptance criteria should be used to evaluate the necessity and sufficiency of spare inventory. If the risk of stockout is lower than broad acceptance region, spares can be reduced in order to release fund occupied by these spares. Otherwise if the risk is bigger than maximum acceptance region, more spares should be stored so as to reduce risk. If the risk is located between maximum acceptable and broadly acceptable region, As Low As Reasonable Practice (ALARP) principle should be followed in spare optimization.

4 Discussion About Risk Acceptance Criteria in Spare Optimization

Unlike tradition engineering risk analysis methodology, economic loss is one of the most important factors that influence the determination of risk acceptance criteria in spare optimization. Human fatality or environment pollution has very few influences on the optimization of spares.

4.1 Analysis of Current Risk Situation

A combined refinery installation in Y Company involves hydrocracking, continuous reforming, FCC and other 6 units. Statistics shown that there are 177 spares belong to Class A category whose unit prices are bigger than 2000 RMB. All these spares are used in important pumps. The type of spares includes rotators, bearings, gears assembly, mechanical seals and other nonmoving parts.

In order to obtain the probability of out of stock, reliability data taken from the Non-electronic Part Reliability Data (NPRD) was used for calculation. All data was checked by engineers from the field. The probability of out of stock was obtained by take into consideration not only spare parts configuration, but also available spare pumps as well. The indirect loss caused by production interruption was quantitatively analyzed. The normal yearly loss (risk) of each spare is illustrated in Fig. 2.

We can see from Fig. 2 that spares with risk range from 0 to10,000 RMB and from 1 to 500,000 RMB per year account for 67 % and 23 % in quantity, respectively. Only 9.1 % of spares own the risk of more than 1 million RMB per year. It seems for those spares whose risks are less than 10,000 RMB per year; there is great room for optimization. While for other 9.1 % high risk spares, prompt spare supplement is essential.

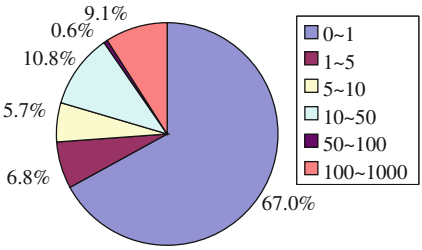


Fig. 2 Normal yearly loss of spares

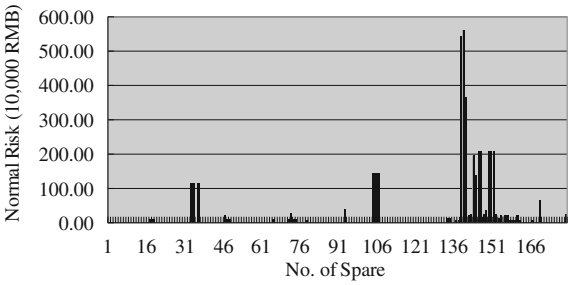


Fig. 3 Distribution of yearly loss of spares

In order to evaluate the risk profile of diverse spares more objectively, the potential yearly loss of all 177 spares are illustrated and compared in Fig. 3. It is shown that there are some spares whose normal risk are much higher than others. Such spares are normally belongs to one or several pumps. The reasons for such risk profile lies in two factors, either the pump is very critical, or there are no spare parts or spare pump available for such pumps. Since the risk is too high to be acceptable, the inventory of such spares should be specially analyzed.

4.2 Discussion About Risk Acceptance Criteria

Theoretically, the risk of spare inventory shall follow normal distribution, which means those spares with very low or very high risks should account for only few proportion. Most spares own medium risks. Suppose that the hypothesis is reasonable, the risk criteria can be determined with the following procedure.

Of all 177 spares mentioned above, it is found that the risks of several spares are too high to be acceptable. According to shell Design and Engineering Practice 《DEP 70.10.90.11-CSPC- SPARE PARTS》, spare whose normal yearly risk is bigger than quarter of its unit price should be stored [9]. If the yearly loss is much higher than its unit price (for example, 10 times), such type of spare should be carefully checked to see whether there is data misuse.

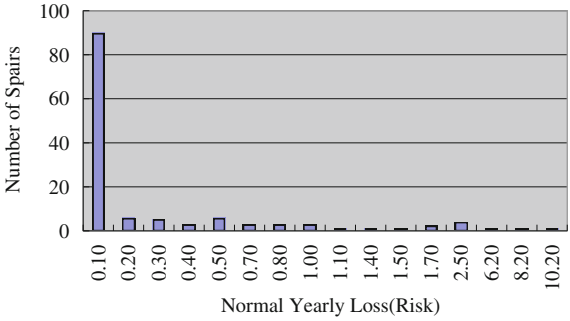


Fig. 4 Normal risk distribution of diverse spares

In our case, 10 times unit price is used as a threshold in data selection. 131 spares are selected for risk analysis and determining risk acceptance criteria. The risk profile of these spares is illustrated in Fig. 4. The risk is range from 0 to 102,000 RMB and the total normal yearly loss due to stockout is 544,3000 RMB.

4.3 Construction of Risk Distribution Formulation

The risk criteria can be determined by the so called equal risk principle. Suppose that the distribution of risks of diverse spares obey the normal distribution, a risk distribution formula can be constructed where the risk of all 131 spares can be covered according to 6σ principle. So the risk distribution formula is constructed as follows:

$$R(x) = \frac{544300}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

Here μ and σ are distribution parameters. $R(x)$ is the total risks of spares whose yearly loss is x. Since all risks of diverse spare are ranged between 0–102,000 RMB per year, so it is easy to obtain the detailed risk distribution formula:

$$R(x) = \frac{544300}{1700\sqrt{2\pi}} e^{-\frac{(x-51000)^2}{5.78\times 10^8}} \tag{2}$$

Provided that the risks of spare do follow normal distribution, it is easy to deduce that those spares with very low risk should be optimized in order to reduce inventory, and others with very high risk should be optimized to reduce risk. Suppose that 80 % of risks own by spares are actually in reasonable region, we can easily determine the risk criteria in spare inventory.

$$\int_{\mu-L}^{\mu+L} R(x)dx = 0.8R_{\text{total}} \tag{3}$$

Here $[\mu - L, \mu + L)$ is the region where 80 % of all total risks were involved. By numerical method it is obtained that,

$$L = 1.28\sigma \tag{4}$$

So the risk acceptance region can be determined,

$$R_{\text{Criteria}} = [\mu - 1.28\sigma, \mu + 1.28\sigma) = [29240, 72760) \tag{5}$$

Here [29240, 72,760) is the tolerable risk region where spare in this region can only be optimized if it meet ALARP principle. Spares with normal risk lower than 29,240 RMB per year is widely acceptable, reduce store inventory is recommended in order to reduce store cost. Spare with normal risk higher than 72,760 RMB per year should be carefully analyzed and prompt supplement might be recommended.

5 Conclusion

Based on the analysis of the characteristics of slow moving spares, the risk of diverse slow moving spares were analyzed. According to statistics and analysis of the risk distribution, the risk acceptance criteria were also discussed in this paper.

(1) A risk based spare inventory optimization method and associated procedure were proposed in this paper. For those spares whose normal risks are less than broad acceptance region, optimizing and reducing spare inventory is recommended so as to reduce fund occupation. While for those high risk spares, prompt spare supplement is essential. For spares whose normal risk range between broad acceptance and maximum tolerable region, ALARP principle should be followed in determination of spare inventory;
(2) A novel equal risk principle was proposed to determine the risk criteria of spare optimization. A normal distribution model was constructed in this paper to study the determination of risk criteria in spare optimization;
(3) The risk of important pump spares in a petrochemical company was analyzed with the method proposed above. According to analysis and statistics, a risk acceptance criteria was also proposed.

Acknowledgments This paper is sponsored by the following projects: The National Key Technology R&D program (2011BAK06B02, 2011BAK06B03, 2012BAK13B03). "863" program (2012AA040103). International Science and Technology Cooperation Program (2010DFB42960).

References

1. van Jaarsveld Willem, Dekker Rommert (2009) Risk-based stock decisions for projects[M]. Econometric Institute, Erasmus University, Rotterdam, p P17
2. Rajeswari, Spare Parts Management [R/OL]. http://productivity.in/knowledgebase/Plant%20Engineering/g.%20Spare%20Parts%20Management.pdf
3. Xuedong CHEN, Bing WANG, Zhibin AI et al (2007) Design and Manufacture of Pressure-bearing equipment based on risk and life. J Press Vessel (in Chinese) 10:1–5
4. Li Y (2000) Current situation of spare parts management in petrochemical industry. Paper submitted for qualification of the senior positions in third training program of sinopec (in Chinese)
5. Brown RG (1959) Statistical forecasting for inventory control. McGraw-Hill, New York
6. Croston JD (1972) Forecasting and stock control for intermittent demands. Oper Res Q 23 (3):289–304
7. Johnston FR, Boylan JE (1996) Forecasting intermittent demand: a comparative evaluations of Croston's method comment. Int J Forecast 12:297–298
8. Yun LUO, Yunxiao FAN, Xiaochun MA (2004) Risk analysis and safety evaluation. Chemical Industry Press, Beijing (in Chinese)
9. Shell Design and Engineering Practice (1995) DEP 70.10.90.11-CSPC- SPARE PARTS[S]. Shell corporation

Architecture Develop Method for Support System of Integrated Joint Operation Based on DODAF

Chaowei Wang, Lin Ma, Qian Wu and Xuhua Liu

Abstract With the gradual development of technology and war concept, Integrated Joint Operation (IJO) has become an important combat mode. The change of operation mode will inevitably lead to change in the development of equipment support and logistics support system (LSS). And the LSS of IJO shows the characteristics that differ from general equipment LSS, such as integration of support resources, networking of command system. Traditional logistics support system developing method appears to be inadequate in dealing with such a complicated system, which gives system description by texts and diagrams. This paper analyzes the functions, organizations and recourses of the IJO logistics support system deeply. Then an architecture develop method is presented with the combination of features of DODAF architecture. The description method of support system is given by DODAF product based on the view products appropriate selection according to the system characteristic. The develop progress and order of system architecture, activities, organization, resource, information and business flows among departments as well as the mapping relationship among the constituent elements is also presented. The results show that the architecture develop method based on DODAF can provide multi-perspective description of the support system, providing auxiliary basis for operational decision making and engineering personnel to improve system performance. The method can contribute effective technical support to the design and establishment of the LSS of IJO.

C. Wang (✉) · L. Ma · Q. Wu · X. Liu
School of Reliability and System Engineering, Beihang University, 100191 Beijing, China
e-mail: wangcwbu@163.com

L. Ma
e-mail: malin@buaa.edu.cn

Q. Wu
e-mail: ustbwuqian@126.com

X. Liu
e-mail: 280175600@qq.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_135

1 Introduction

An integrated joint operation is the major combat mode under conditions of information age in the future. Much importance has been attached to IJO by many military institutions military and it is increasingly concerned by a large number of counties [1]. Integrated joint operations puts emphasis on the confrontation and the overall contest between combat systems rather than the single equipment, The integration complementarities and dependence among the various armed forces including space force, electromagnetic force and intelligence force, will be enhanced, which put forward higher requirements for the overall function of the combat system. Equipment logistics support system, as an important part of combat system, has to form coordination with the combat system. The system should improve logistics support capabilities and develop new logistics support mode to match the operation system, which must provide powerful integrated and comprehensive logistics support for combat equipment foe combat ordnance equipment. Logistics support capability under integrated joint operations must be able to adapt to multidimensional integrated battle space of joint operations. Only in that way can equipment maintain excellent performance to win the war through logistics support activities.

The logistics support system of integrated joint operations shows a variety of characteristics, which are different from the traditional support system, such as many factors involved, varied tasks, complex structures and organizations, onerous business and numerous information flows [2]. The design and development method of traditional system, which utilizes texts, forms and graphics to describe system, cannot meet the new requirements. The application of standardized system architecture design model in the development of logistics support system of IJO is imperative. DODAF, as an increasing popular system architecture design tool, is able to promote a better understanding of requirements, and provides clearer design method and more convenient maintenance means. Simultaneously with the relationship among views products of DODAF, accurate system modelling and design control can be feasible for the developers. In addition, DODAF products also offer the function of vivification. The identification of requirements the validation logic behaviours and system design integrity can be checked automatically as early as possible.

2 Application of DODAF in System Architecture Development

2.1 Introduction of DODAF

The Department of Defense Architecture Framework (DoDAF) is an architecture framework that provides structures for a specific stakeholder concern [3]. It is developed by the United States Department of Defense (DoD). DODAF consists of a series of views products, which can be summarized as all views(AV),operational

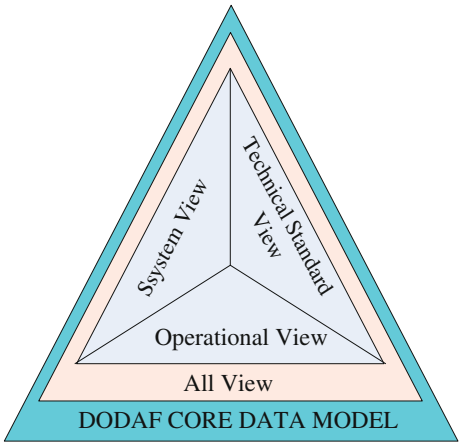


Fig. 1 DODAF view products

views (OV), system views (SV) and technical standards views (TV), shown as Fig. 1. These views act as mechanisms for visualizing, understanding, and assimilating the broad scope and complexities of an architecture description through tabular, structural, behavioral, ontological, pictorial, temporal or graphical means. DODAF is especially suited to large systems with complex integration and interoperability challenges, such as system of systems (SoS), family of system (FoS).And DODAF is apparently unique in its use of "operational views" detailing the external customer's operating domain in which the developing system will operate [4].

2.2 Development Progress of DODAF Products

As the complexity of the system itself, to develop system architecture with certain features through architecture framework (DODAF) is also a very complex process. To establish system architecture with DODAF view products, develop progress of products should be determined. As shown in Fig. 2, the develop progress includes the following steps.

- To determine the intended use of the system architecture

In most cases, developers of system architecture have no enough time, funding or resources to build a all-encompassing system model. Therefore, the system model should focus on a specific target. Before describing the system architecture, developer who is responsible for designing should clarify and confirm the requirements, expectations objects and application prospects of the system.

- To determine the scope, background, environment, and other conditions of the architecture

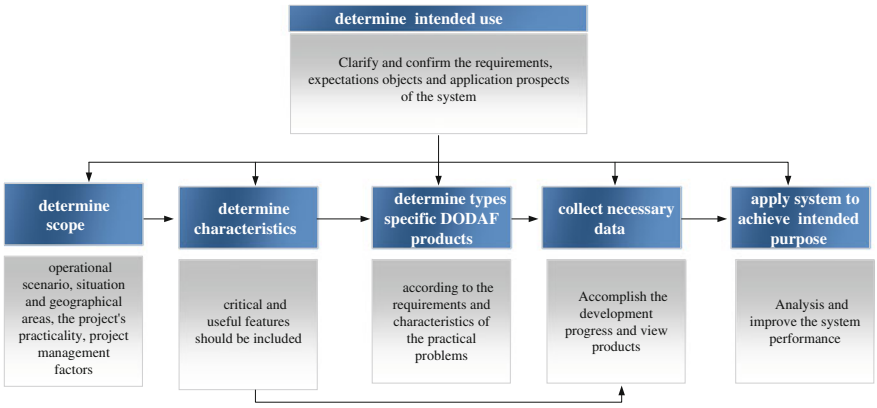


Fig. 2 Develop progress of DODAF

Once the purpose and use of system architecture has been determined, many issues should be taken into consideration. The issues consists of the scope of the architecture (behavior, function, organization, time, etc.), the appropriate detail of the architecture in a wider range of application background, the operational scenario, situation and geographical areas, the project's practicality, project management factors and other relevant factors.

- To determine the characteristics of the system architecture

According to the object and purpose of system architecture, what features of the system need to describe t should be careful studied. If critical and useful features are ignored, the final architecture may be useless and cannot archive the determined goal. Also, if it contains a lot of unnecessary features, the developed architecture would be viability under the conditions of a given time and permitted available resources. In addition, the development of the architecture should be sufficiently forward-looking to make the system be able to adapt to reuse and expand in future.

- To determine the types, specific products and establish order of views in DODAF for the system architecture

In many cases, it is not necessary to create all of the given views products in the framework Developers should make a appropriate selection in accordance with the requirements and characteristics of the practical problems.

- To collect necessary data of system architecture and to establish the view products

The chosen view products in DODAF would be created with the collection, correction and combination of the necessary system configuration data. The system architecture described from different viewpoints requires a variety of products.

At all stages of system development, correlation and consistency the view products should be verified under the permitted conditions.

- To apply the system to achieve the intended purpose and to analyze and improve the system performance

The system architecture will be put into use for combat or business process restructuring and reorganization processes, to improve operational efficiency, and information systems planning. And it can provide documentation requirements for mangers and purpose suggestions for improving system performance. System architecture established by DODAF can provide support to achieve the above object, but it itself cannot offer the actual conclusion or answer. So conclusions must be drawn from the analysis with the combination of the actual usage.

3 Logistic Support System of IJO

3.1 Integrated Joint Operation and Its Logistics Support

Integrated joint operation is multidimensional space combat military actions, which focus on a unified objective of the operation. And the combat units or elements of system integrate in order to perform better. Integrated joint operation put emphasis on the confrontation among different combat system rather than single combat units or combat elements, which is different from conventional combat mode. Thus the outcome of combat primarily depends on the overall performance. The change of the combat mode have a broad and profound impact on the equipment logistic support and put forward newer and higher requirements on the capability of logistic support [5].

Logistic support of IJO stresses the systemic and holistic logistical support of the whole combat system other than single equipment. Focusing on the requirements of logistic support, the support resource, units, organizations and other elements of various armed force should make comprehensive integration, which includes navy, air force, space, and eclectic. The formation of logistic support needs to adapt to multi-factor optimization. The combination of the three major elements, as known material, energy and information, must be efficient and stable so as to provide continuous, secure, and efficient logistical support for the war.

3.2 Characteristic of Logistics Support System

The equipment logistic support system is the integration of all the necessary support recourses and management factors, which use for the operation and maintenance of the ordnance equipment. Logistic support system is combination of interconnected

elements to constitute an organic whole entity, which consists of three kinds of elements, support functions, support organizations and support resources, shown in Fig. 3. The function of support system is to maintain the normal use of equipment. It ensures that support system is able to accomplish different tasks when the equipments are in usage or in failure. To achieve specific support system functions, specific support activities must be carried out in the appropriate support organization. The implementation of these activities requires usage or consume of support resources. The appropriate support resources should deployed to the appropriate support organization, in order to assure the normal execution of the logistic support system function. Integrated joint operations have the features of vast battlefield, the huge material consumption, frequent military maneuver and accelerate the pace of operations. The various change of equipment tasks lead to the frequent change of support tasks. The structure of many armaments is more complex, with high automation and intelligence. The complexity of the equipment protection system put forward higher requirements on the logistic support system. The support system is composed of people, equipment, facilities, spare parts, technical data and other properties, which may be located in different regions. The links among these huge elements and the management factors are complex, which brings the inherent complexity of the systems. The following analysis of logistic support system will focus on the support factions, support organizations and support recourses.

- Support factions

Integrated joint operations equipment tasks are very complicated, and the risk of equipment battle damage greatly increases. Compared with general equipment logistics support systems, logistics support system of IJO should focus on strengthening battlefield repair capabilities as well as functions of corrective maintenance and preventive maintenance. Meanwhile, the support system should also have the function of investigation, intelligence gathering capabilities and decision-making skills so as to be effective access to real-time battlefield environment and equipment information to quickly identify equipment support programs, organization support forces to reach the designated area. In addition, the integrated joint operations battlefield environment is even worse. Support systems should also have anti-strike capability for its own survival.

- Support organizations

The organization of integrated joint operational support system should be integrated and form a network. Different establishment, different units, different levels with similar functions should get together to form cross-cross, stable and reliable integrated security network. The integration of network can make each agency and units can get the location of combat troops and its support needs, and let the logistical support personnel to make forward-looking arrangement in accordance with the requirements of time and place of combat troops to provide the necessary support. The support personnel would use the information network systems, to collect support needs, analyze support situation, and pay attention to the support

process. And they have to ensure smooth delivery and use of network resource allocation to improve the overall security performance.

- Support recourses

Integrated joint operation logistics support system put emphasis on efficient support. The system uses knowledge collaboration such as establishing intelligence and knowledge center and other forms of support, knowledge and intelligence to play a role in make out the support demand for real-time and scientific, rational and effective forecasting for all types of support resources. Means of combat support requirements of different forecasting and timely amendment ensure that resources in terms of time to achieve optimal allocation. A variety of support means, such as remote support systems, interactive electronic technical manuals (IETM), integrated test equipment make the logistics activities completed quickly and efficiently.

4 Logistic Support System Architecture Development Based on DODAF

4.1 ABM Based on DODAF

The second chapter shows that DoDAF consists of 26 view products from different viewpoints to describe the architecture. But in the actual development process, the design and development of so many products will bring burdensome for the designer. According to the needs of different users, several view can meet certain needs. So the developers do not need to use all views. There is no standard for selection of DODAF products, which brings the system design and develops staff a lot of difficulties. With the U.S. Department of Defense Architecture Framework developing, the researchers summarized the develop progress of DoDAF products. Steven J. Ring and other researchers proposed activity-based modelling approach to solve the above problem.

ABM method is a three-view modelling approach, which is different from the previous product-centric design and analysis methods. It emphasizes data of the architecture products based on strict rules, supporting cross among different products. The method provides ways to automatically generate a part of architecture products. The main characteristic of ABM method is operational views have corresponding relation with the system view. Architecture can be divided into three objects: entities, relationships and attributes. Entity is the objects which process and store architecture data; relationship is the links between entities; attributes belongs to entities and relationships, are characteristics of entities and relationships.

It can ensure the consistency of data in the design process that DoDAF views products are developed with ABM method. In operational views, information exchange and demand line can be automatically generated by the software after activity; mechanisms, information input and output in OV-5 are created manually.

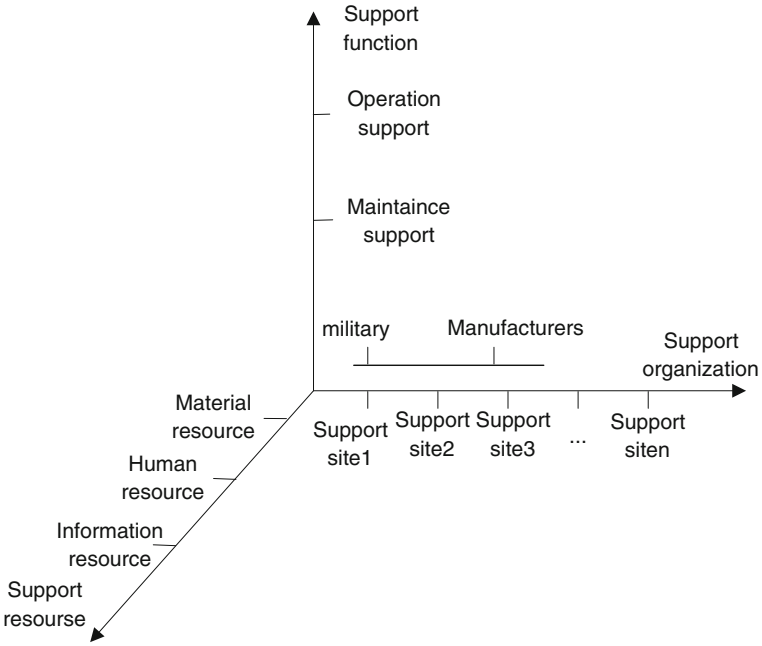


Fig. 3 Elements of logistic support system

Meanwhile, when all of the information exchange within the portfolio is complete, information exchange matrix OV-3 can automatically be generated. Similarly, the system views, the system data exchange and interfaces in SV-1 can be automatically generated after mechanisms, data input and output in SV-1 are created manually. SV-6 can be automatically generated, too. The relationship among them is shown in Fig. 4 [6].

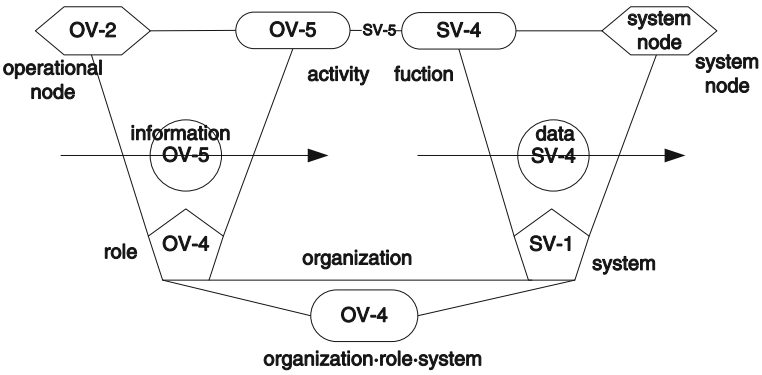


Fig. 4 Relationship among views in BM method

4.2 System Architecture Develop Progress Based on DODAF

The modelling progress of integrated joint operations logistical support system is the progress of mapping from integrated joint operations logistical support activities to the system functions process. First, establish OV-1, which makes system designers have a complete understanding of the architecture. OV-5 is the core view products, with the operational node connections described OV-2 and organization described in OV-4. Similarly, in system views, SV-4 is the core view products. SV-1 and SV-3 can be made out easily based on SV-4. Meanwhile, SV-5 is the link between operational views and system views, mapping entities, relationships and attributes in operational views to the system view. The modelling process is shown in Table 1.

Because operational views and system views are correspondence and modelling methods and procedures of the two types of DODAF products are basically the same, so only operational view modelling specific steps are studied as follows:

- Create OV-1 of logistics support system of IJO

OV-1 describes the mission major system nodes, and the interaction between each node from a macro perspective intuitively.

- Establish OV-5 of logistics support system of IJO

Activities model describes activities executed to complete a task or series of tasks, including the activities of the task logical relationship among activities and tasks as well as the input-output relationship between environment and activities. Establish OV-5 is mainly divided into two steps, namely to establish a tree diagram and exploded the views layer by layer.

- Establish operational node in OV-2 of IJO

OV-2 describes the nodes and the exchange of information among nodes of IJO support systems. It also reveals the node entities and their duties. When establish

Table 1 Modelling Progress based on ABM

Development of Operational Views	Development of System Views
Create operational activity model	Create operational functionality model
Create operational node in OV-2	Create systems/services interface description
Create organization in OV-4	Create systems-systems/services matrices
Create relation among activity, operational node and organization	Create relation among functionality, node and system
Generate relationship among activity, operational node and organization	Generate relationship among functionality, node and system
Create operational information exchange matrix	Create systems data exchange matrix
Generate need line among operational node	Generate links in systems interface data

OV-2 manually, only need to define the operational node, including internal nodes and external nodes. For a complete OV-2, it also needs to reflect the needs of lines between nodes, which can be generated by the software automatically.

- Establish organization's role and organizational units in OV-4 of IJO

OV-4 describes structural relationship among the internal architecture of the entity, the entity type and the external entities. The types of relationship includes: management reporting relationships, command and control relationships, obey relationship and collaborative relationships. To establish an organizational diagram can clearly explain the relationships in the structures. And the relationships between the activities and organization are closely linked.

- Create relation among activity, operational node and organization manually of IJO

It is needed to Create Relation among activity, operational node and organization manually.Activity, operational node and organization comes from OV-5, OV-2 and OV-4 established. The ternary relation is not one to one, an activity can correspond to multiple nodes, and can also correspond to roles,

- Generate relationship among activity, operational node and organization automatically of IJO

After manually create a correlation matrix, the system has automatically generate, and ternary relationships among activities, nodes and roles. At this time, activity, operational node and organization is automatically related in OV-2, OV-4 and OV-5.

- Create operational information exchange matrix of IJO

DODAF development tools can automatically create operational information exchange matrix. The DODAF development tools can also automatically check the validity and accuracy requirements.

- Generate need line among operational node of IJO

The information exchange among operational nodes in OV-2 can be generated automatically by the DODAF development tools.Thus the operational node connectivity description (OV-2) is completed.

After completion of the above 8 steps, the integrated joint operation logistic support system model is accomplished. The model based on DODAF makes a comprehensive description of the system equipment support system mission objectives, equipment information, and activity information and information exchange. On this basis, this model can also be used to establish the corresponding simulation model for specific support tasks.

5 Conclusion

This chapter analyzes characteristic of logistics support system of integrated joint operation deeply. And an architecture develop method is presented with the combination of features of DODAF architecture is presented. The results show that the architecture develop method based on DODAF can provide multi-perspective description of the support system,providing auxiliary basis for operational decision making and engineering personnel to improve system performance.On the basis of the model built by DODAF products, model for simulation can be easily built to test and verify the design of business and information in the system architecture, which can improve the performance of the system. The develop method can contribute effective technical support to the design and establishment of the LSS of IJO.

References

1. Mutschler DW (2006) Enhancement of memory pools toward a multi-threaded implementation of the joint integrated mission model (JIMM). In: Proceedings of the 2006 winter simulation conference
2. Alex G (2008) System-of-systems engineering management: a review of modern history and a path forward. Syst J 2:484–490
3. DoD Architecture framework working group (2004) DoD architecture framework version 1.0. USA, 2004-01
4. DoD (2001) Systems engineering guide for systems of systems 2008. In: Gibbs JT, Huang LN (eds). VSA DOD, pp. 7–10, 16–23
5. AR 700-127 (2007) Integrated logistics support. HQ USAF, Washington DC
6. Pan X, Yuanxing H, Baoshi Y (2012) Equipment of system of systems architecture based on function and connection. Syst Eng Electron 34:2052–2057

Research on the Training of the UAV Operators

Tian Yong, Zhang Wenjin, Yang Xinglei and Wen Yu

Abstract Compared with manned aircraft, Unmanned Aerial Vehicle (UAV) has small volume, low cost, convenient use, low requirements for combat environment, and strong survival capacity in battlefields. UAV has become the focus in various countries armies. All countries in the world are actively developing UAV, but for UAV operator training, both theoretical research and practice is still exist many deficiencies. This article starts with classification, task model and characteristics of UAV, then analysis the operators' requirements of different task models, and researches UAV operators' training of Chinese army base on using foreign army's experience as references, finally provides reasonable suggestions.

1 Foreword

Unmanned Aerial Vehicle, which was called UAV, was a kind of aircraft that did not have driver in machine, relying on the power drive, and controlled by the vehicle interior automatic control system or remote control command by external control station transmits. It was mainly used in Battlefield reconnaissance, damage assessment, fire attack, communication monitor, electronic interference, long-range fire, spotting and so on. It was an important means to improve the ability of long-range precision strike, ability of information warfare and electronic warfare capabilities.

Compared with manned aircraft, UAV was smaller in size, lower cost, more convenient to use, lower requirements for battle environment and stronger battlefield survivability. UAV had gotten the focused attention of armies of the whole world and also obtained the rapid development since the beginning of twenty-first century.

T. Yong (✉) · Z. Wenjin · Y. Xinglei · W. Yu
School of Reliability and Systems Engineering, Beihang University,
The Unit 95927 of PLA, Beijing, China
e-mail: ty@dse.buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_136

2 Classification of UAV

The Air Force of the United States announced The US Air Force UAV System Flight Plan 2009–2047 in 2009 in order to provide guidance for future UAV system development. This plan classified the UAV as large-size, medium-size and small-size. Large-size UAV systems focused on the development of RQ-4A/B Global Hawk; medium-size UAV systems focused on the development of MQ-1 Predator and MQ-9 Reaper; and small UAV systems focused on the development of Wasp III, RQ-11B Raven and Scan Eagle [4]. The comparison of three types of UAV data (The Official Website of the US Air Force 2013) was shown in Table 1.

Due to the difference between large/medium-size and small-size on take-off weight, flight altitude, endurance and payload, three types of UAV had completely different task models (Table 2).

3 Task Model and Characteristics

3.1 Task Model and Characteristics of Small-Size UAV

Due to the low ceiling, slow speed and short endurance, the main function of small-size UAV was reconnaissance, usually provided tactical level support for the ground force. For example, RQ-11B Raven was mainly used for battlefield situational awareness, target acquisition, damage assessment and ground threaten detecting evaluation for the platoon troop. In the following we took RQ-11B Raven as an example to illustrate small-size UAV task model and the characteristics.

Typical task model: the combatants in a platoon launched small-size UAV by approach, hand thrown and catapulted way; and then the operator remote controlled the UAV to investigate the battlefield environment or monitor the important targets in the third perspective.

Figure 1 had shown that Sgt. Dane Phelps, from 2nd Battalion, 27th Infantry Regiment, 25th Infantry Division prepared to launch the Raven unmanned aerial vehicle during a joint U.S. and Iraqi cordon and search operation in Patika Province, Iraq [16].

By analysing the task models, we found out that small-size UAV had following characteristics when completing the pre-planned missions (Fig. 2).

(1) The UAV operators controlled with the third perspective during the whole mission time, which was similar with the remote controlling of aircraft model;
(2) Distance of data link between operators and UAV was near, and it was using line-of-sight mode as communication model;
(3) UAV operators were usually based personnel in the frontline combat units.

Table 1 Comparison of performance between UAV

Specifications	Large-size UAV	Medium-size UAV		Small UAV		
	RQ-4 Global Hawk	MQ-9 Reaper	MQ-1B Predator	Scan Eagle	RQ-11B Raven	Wasp III
weight (kg)	14,628	4,760	1,020	18	1.9	0.453
Ceiling/Operating altitude (m)	18,288	15,240	7,620	4,876	152	152
Speed (km/h)	574	482	217	88 ~ 128	48 ~ 96	32 ~ 64
Payload	electro-optical, infrared, SAR, and high and low band SIGINT sensors	AGM-114 GBU-12 GBU-38	AGM-114 AIM-92	High resolution, day/night camera and thermal imager	High-resolution, day/night camera and thermal imager	High resolution, day/night camera
Range(km)	16,112	1,850	1,239	30	8 ~ 12	5
Endurance(hours)	>28	14	24	>20	1 ~ 1.5	0.75

Table 2 Training course of UAV operators [18]

Student type	Training content				
UAS operators	Initial flight training: 2 months	Qualification course of UAS control: 2 months (include 46 h simulator training)	Basic course of UAS: 1 month (include tactical and battlefield operation: weapon, radar, sensor, threat analysis; 100 h theory training; 7 times' airborne laser detection battle missions)	Course of joint fire	Refit training of MQ-1 force: 3 months
Sensor operators	Basic course of aircrew: 3 weeks	Basic course of sensor operators: 5 weeks (include full motion video training, sensor basic course, intelligence analysis)		–	

Fig. 1 A soldier prepares to launch the Raven in Iraq [16]

3.2 Task Model and Characteristics of Large and Medium-Sized UAV

Large and medium-sized UAV had high flight speed, long endurance, large radius in scope of flight, and large take-off weight. They could configure a large number of optical, infrared, synthetic aperture radar and other electronic equipment to complete strategic or operational level intelligence reconnaissance, communication relay, battle management control and other tasks. Some UAV had all the capability of both intelligence reconnaissance and destruction. For example, MQ-9 UAV primarily supplied for ground force of brigade level, and they also provided a

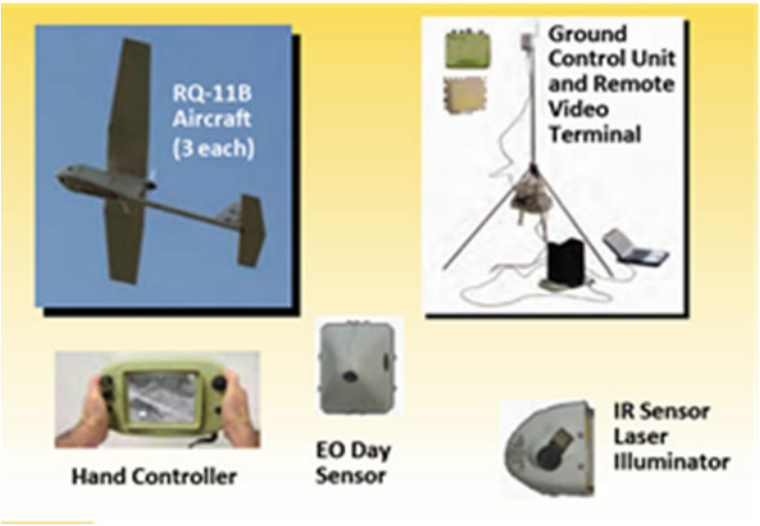


Fig. 2 The component of *RQ-11* [16]

variety of payloads and strike capabilities, supported to complete tactical and operational levels of combat missions. In addition, Reapers also performed the following missions and tasks: intelligence, surveillance, reconnaissance, close air support, combat search and rescue, precision strike, buddy-laser, convoy/raid over watch, route clearance, target development, and terminal air guidance.

The next we took MQ-1B predator as an example to illustrate medium-sized UAV task model and characteristics.

The command centre of division or brigade level received fire support request from frontline soldiers and completed planning mission according to combat missions. The frontline battlefield operators (field operator) launched the UAV and took-off by ground rolling from the ground control station that installed in frontline. After lift-off, the UAV flied in the panning route and handed over control to operators (Infield operator) in field ground station, and Infield operator remote controlled the UAV by data link communications. Upon the completion of reconnaissance, positioning and locked on, then UAV attacked the enemy target, and assessed the effect of the damage, subsequently returned following the panned route, and finally landed under the control of the field operators.

Each Predator UAV was equipped with a system operator ("Pilot") and a sensor operator, and they controlled the Predator UAV in the ground control station.

As Fig. 3 revealed, Captain Richard Koll, left, and Airman 1st Class Mike Eulo perform function checks after launching an MQ-1 Predator unmanned aerial vehicle August 7 at Balad Air Base, Iraq. Captain Koll, the pilot, and Airman Eulo, the sensor operator, will handle the Predator in a radius of approximately 25 miles around the base before handing it off to personnel stationed in the United States to

Fig. 3 Predator operators at Balad Camp Anaconda, Iraq, August 2007

continue its mission. Both are assigned to the 46th Expeditionary Reconnaissance Squadron [2].

Task characteristics of Large and medium-sized UAV:

(1) Medium-sized UAV had long flight mileage, and long endurance, so it was usually used as "remote control division [19]" mode. The field operators took off and landed the UAV, and the infield operator was responsible to avoid the danger zoon, searched targets, delivered weapons, damage assessment and other tasks.
(2) Either infield or outfield operators manipulated the UAV in the ground control station in the first person perspective through the video from UAV returned.
(3) Because the long distance, normally more than 200 km, from the UAV to the infield ground control station, operators often used satellite data transmission or UAV relay communication to command UAV activities.

A notable characteristic of UAV was "Self-service machine, but manual-control system." Currently the capability of UAV autonomy was still quite low. In 2002, the Air Force Research Laboratory (AFRL) [4] established autonomous control level (Autonomous Control Level, ACL) of UAV system in accordance with the OODA model, which had total of 11 levels. With enhance of UAV autonomous ability, UAV systems was classified from a remotely piloted (0) to a fully autonomy (10), and corresponded different requirements of perception, cognition, analysis, coordination, planning and decision-making, mission capabilities and other aspects accordingly. In aspect of UAV system, Predator (MQ-1), Global Hawk (RQ-4) of U.S. military had achieved 2–3 ACL. Joint Unmanned Combat Air System (J-UCAS) and X47-B would achieve 5–6 ACL [18]. Therefore currently, in order to complete the task, the operators mainly controlled the UAV through the control loop. Different types of UAV had different task models and characteristics, so the UAV operators' competency requirements were also different.

4 Analysis of UAV Operators' Capacity Needs

4.1 Analysis of Small-Size UAV Operators' Capacity Needs

Small-size UAV systems were cheap prices, low altitude and short endurance, so they always performed tasks such as monitoring and detections. Generally grass-roots combatants controlled UAV by using the third person perspective mode, and this approach was similar to the ground remote control flying model. Therefore, in terms of small-size UAV operators, the manned driving experience was not required, but practical operating skills were needed, especially during take-off and landing phases. The sensor operators who cooperated with the former needed to have a good synergistic capabilities, intelligence reconnaissance and analysis capabilities to jointly accomplish tasks. For example, soldiers operating RQ-11B "Raven" UAV only needed to complete a 10-day total of 80 h of theoretical and operational instruction courses [14].

4.2 Analysis of Large and Medium-Sized UAV Capacity Needs

For large and medium-sized UAV, the price, flight altitude, endurance, and type of job performed complexity and importance were all different that small-size could not be compared. So with the change of operational target and requirements, there were higher requirements for the operators on skills and overall quality, it was necessary to master the operation control technology, and should have a decision making capability.

Specifically, large or medium-sized UAV endurance was usually ten hours or even tens of hours, so large and medium-sized UAV commonly use "Remote Control division" approach by field operators and one or more infield operator groups in turns. Using the remote control division, the field operators were only responsible for took off and landed, on the other hand infield operators were responsible for target searching, weapon delivery, damage assessment and other tasks by autopilot controlling to enter and exit the theater. This division labor method could abate the requirements of the operators' quality, and could also reduce nearly one-third of training time [19].

(1) Field operators' competency requirements

Using "Remote control division" method, the main responsibility of field operators were taking off and landing UAV. Taking-off and landing phases, which were considered to be the most difficult periods and the most critical phases of UAV mission, required operators' high standard flight experiences. During these phases, field operators manipulated through the screen video that UAV took back from the

controller. That was why field operators were required to have some flying experience and skilled ability to take off and land UAV, to percept UAV three-dimensional situation through a two-dimensional video that was displayed by the ground control station.

(2) Infield operator (Pilot or UAV system operator) competency requirements

- Using "Remote control division" method, infield operators were not required to take off and land UAV. After entering the war zone, however, they still need complete manual controlling to avoid hazardous areas, target acquisition, weapons delivery, damage assessment and other tasks. So infield operators were also required to have some flight experiences that permitted UAV three-dimensional situation through a two-dimensional video that was displayed by the ground control station. In addition, for ensuring the completion of the task, UAV control station must have the ability to operate the equipment of UAV.
- When large and medium-sized UAV were performing remote tasks, the operators controlled UAV through remote data linked by satellite, so there was a delay on the images and videos that UAV returned. Operators manipulated UAV according to the delaying pictures might make a wrong decision to UAV attitude, and then overkilled the subsequent amendments because of the delay, and induced oscillation as a result, finally would put UAV into hazard state. It required that operators should have strong ability to predict in order to avoid accidents, according to the operation of the steering column to predict UAV attitude.
- Conditions' changing in battlefield were rapidly, it required operators had strong information integration, intelligence analysis and decision making ability by manipulating UAV through the images and video signal UAV returned, and made analysis and decision quickly and accurately according to battlefield, target location, path, etc.

(3) Infield sensor operators' competency requirements
Sensor operators needed to assist infield operators to complete weapon delivery, damage assessment and other tasks through optical, infrared, radar and other equipment on UAV. Thus sensor operators should be:

- Able to use a variety of optical, infrared and other sensors configured on UAV skillfully, and had strong map analysis and intelligence analysis ability.
- Had good collaboration ability that helped infield operators' complete scheduled tasks.

5 Training of U.S. UAV Operators

The training system, which is specialized and normalized, is very important for combat capability's generating of UAV system. And the most important part is the mechanism of personnel selection and the method of training.

5.1 Personnel Selection

The report of U.S. Air Force [7], Under Secretary of Defense for Acquisition, Technology and Logistics, [15] indicated that operators of Predator would be trained specialized for their accreditation, experience of manned aircraft was critical for Predator's accurately control, and it contribute to UAC's operation.

U.S. Air Force considered that the ability of large UAV's operators is nearly the same to the skill of manned aircraft's pilots. The reason was those pilots who were skillful and perceptive have enough experience for reducing risks. These risks were likely to occur when UAV aviated on complex and crowded airspace, and when dropped bombs to the place closed to own side.

In the early time of Predator's development, accordingly, US Air Force was inclined to demand temporary training from experienced pilots, who would be practiced to UAV operators after training. This kind of measure ensured the success of Predator in the early time. However as UAV's development and mission's increasing, the measure's defects were revealed gradually. Firstly, the effects of payment and promotion became the most prominent problem. Highly qualified pilots were unwilling to transfer their posts to UAV operators that resulted to serious shortage of UAV operators. Secondly, highly qualified pilots spent many weeks or more for UAV adaptive before performed UAV missions. When missions completed, it raised task cycle and costs as a result of long time grounded and recovery training. In addition, the measure of temporary training had no practical contribution to solute the predicaments of operators' shortage. It was still lack of manned aircraft's pilots and UAV operators in condition of large scale war.

Compared to Air Force, U.S. Army had no requests of flight experience for UAV operators. The main reason was insufficiency of pilots.

"My concern is that our services are still not moving aggressively in wartime to provide resources needed now on the battlefield. I've been wrestling for months to get more intelligence, surveil-lance, and reconnaissance assets into the theater. Because people were stuck in old ways of doing business, it's been like pulling teeth.… All this may require rethinking long-standing service assumptions and priorities about which missions require certified pilots and which do not." [2] Referred by Robert.M.Gates on April 2008, who was the former U.S. secretary of defence, questioned that future unmanned aerial vehicle system (UAS) operators' requirement is the certification of high qualified pilots.

According to General Richard Hawley, former commander of Air Combat Command, "I've spent time in a UAS control van. You don't need 500 h of F-16 time to know how to fly a Predator. You do need to understand something about winds, weather, and the environment in which the Predator operates." [2]

Therefore the major reform was executed which aimed to the training of U.S. UAV operators. For the purpose of better UAV operators' cultivation of U.S. Air Force, declared by Michelle Morris on February 2008, the U.S. CSAF, a school of UAV weapon would be established in Nellis Air Force Base in July. The main goal of teaching was educating specialized UAS operators, sensor operators and ground maintenance personnel that replaced current measure that transferred pilots temporarily from Air Force. In September 2008, the CSAF ordered that employing new operators for UAV control, and meant that the new occupational areas that test admit and comprehensive training for select UAV operators were set up [18].

5.2 Training Content and Method

The major of UAS student of U.S. Air Force divided into 2 parts—UAS operators and sensor operators. The training content included basic course of aircrew, basic course of sensor operators, qualification course of UAS control and basic course of UAS.

The congress report that submitted by Defense Department pointed out that the shortage was still severe either UAV operators or sensor operators in current [15].

In the following Table 3, the manpower requirements were presented on December 16th 2011, which was for Remote Piloted Aircraft (RPA) pilots and Sensor Operators (SO) to support 57 MQ-1/9 and 4 RQ-4 Combat Air Patrols (CAPS). It included operational, test, and training requirements, as well as appropriate overhead and staff requirements.

By comparison, presented in Table 4, the practical number of trained RPA pilots and SO available on December 16th 2011 made clear that the personnel shortfall could not be ignored.

Table 3 RPA crew manpower requirements

	MQ-1	MQ-9	RQ-4	Total
Pilots	1,012	529	155	1,696
SO	730	401	63	1,194

Table 4 Current RPA crew manning availability

	MQ-1	MQ-9	RQ-4	Total	Current Shortfall
Pilots	726	455	177	1,358	−338
SO	610	291	48	949	−245

From Tables 3 and 4 we known that a large number of missions were executed in worldwide by MQ-1, MQ-9 Predator, therefore the need of UAV operators were urgent strongly, and the number shortage was obvious. For high costs, it was impractical that training operators depended on UAV totally. So the high degree simulator for UAV operators training was considered, the number and quality of training personnel increased at the same time that the costs and cycle times were saved by this measure. This view was also evidenced in report [1].

In order to replied the rapidly development of UAS and met the need of UAS training scale quality, U.S. Army was cultivating the training ability of UAS, the main measure were list following.

(1) Researching and developing advanced UAV training system and distributed and multi-services UAS training. Prepared the new challenge that Army was operating with mixed formation of manned aircraft and UAV, or performing coordinated with multi-UAV in the future.
(2) Besides developing advanced UAV simulator, the ground control station was also improved. The Block 50, new constructed ground control station of cockpit, was designed for giving consideration to both operation and training. In a view of human factors engineering, the simulated training environment would be more authentic, and the quality of training would be more improved.
(3) Furthermore, the training mode and method of UAV operators was strengthened by U.S. Army. It was mentioned in reference [6] that Functional Near-Infrared Spectroscopy (FNIR) applied to optical brain image technique for detecting the response feeling of brain action. By this method, the training and estimation of UAV operators would be enhanced effectively. Moreover, Ref. [5] indicated that a new approach which followed studies on human factors performance and cognitive loading. The resulting design serves as a test bed to study UAV pilot performance, create training programs, and ultimately a platform to decrease UAV accidents. As a result, the UAV driving system with integrated motion cueing was developed, that was used to training and estimating UAV operators.

6 Suggestion of Chinese Military for UAV Operators Training

6.1 Drawing up a Long-Term UAV Plan

Heading As mastering the most powerful UAV technique, the Air and Land Force of U.S. had drawn their UAV plan relevantly [3, 4]. In order to confront the challenge of new revolution in military affairs, Chinese Army should set up a long-term plan for development of UAV and cultivation of operators, which aimed to better direction of UAS.

For instance, current UAV was controlled by operators who used data link. It went without saying that data link was very important to UAV. In future battle field, our army would operate with mixed formation of manned aircraft and UAV, or performing coordinated with multi-UAV. In order to ensure the data link unblocked and available, it was necessary to plan radio spectrum resources of data link.

In addition, the operation mode of ground control station need to be planned permanently. Operation mode of UAV that were same types should be normative and unified so that better replied the challenge that multiple UAVs were controlled by one ground station, or controlled by one operator in the future battle field.

6.2 Solving the Problem of UAV Training Formation and Structure

Recent year, with the development of UAS, UAV was appearing in every field of battle space. In modern warfare, it played more and more critical roles by performing tasks even some complex missions that manned aircrafts were hard to achieve. Some experts referred that the modern battle style was changed by UAV, and it would probably replace manned aircraft's position. Experience had shown that the difference between UAV and manned aircraft were larger and larger [2].

Consequently, from a long term trend, Chinese Military had certainly needed to establish academy which trained UAV operators and maintenance personnel specially. Lessons learned from U.S. Army had proved that talent of UAS would be developed sustainably only by establishing professional UAV troop to promote operators or maintenance personnel.

6.3 Selecting Talents

In order to avoid repeating the faults of U.S. Army, selecting scale of our UAV operators should be broaden reasonably. From the situation of our army, the national defense students of aviation academy and former pilots who were grounded due to physical status were selected to be UAV operators suitably.

6.4 Training Courses and Methods

In regard to large and medium size UAV, training of UAV operator could be relied on aviation academy by applying 2 + 2 methods. In academy, cultivated students accomplished some basic courses for 2 years such as principle of flights, aerodynamics, and automatic control theory. Then these students would be arranged to finish the initial flight training, equipment operating qualification test, and tactical

operating training. The method relied mainly on theoretical and simulation training while making flight training and UAV modification subsidiary.

Generally, on the other hand, the operators of small size UAV could be selected from outstanding sergeants of grassroots units and trained short term.

After operators were selected into UAS troop, the subsequent supplementary training should be put into practice by modular way which was according to mission requirement. In addition, UAV operators' flight training should be regular that could keep perception to UAV attitude.

6.5 Enhancing the Analysis of Training Equipment and Measure

Professionalism, Specialism, and normalization were critical to UAV operators' training effectively. In the future, the training would be relied on multiple types of UAV simulative equipment. Therefore the fidelity of simulative equipment was also critical to training quality. Besides, UAV ground control station that included land, aviation, and navy should give consideration to both battle and training in design.

Moreover, experience of U.S. Army could be used for reference, which was leading medical method such as psychological assessment and behavioral pattern analysis into operators training and estimation. By this method we could solve training effect evaluation better, and avoid the deficiency of evaluation only depended on test and questionnaire.

7 Conclusions

In pace with the development of technology, UAV had played more and more significant role in modern warfare. Operators were kernel of UAS. In order to better adapt the development of UAS, a long term plan should be established reasonably. As developing UAV equipment rapidly, operators training should also be paid attention at the same time, and related research of training method and effective evaluation need to be done.

References

1. Cambone SA (2005) Unmanned aircraft systems roadmap 2005–2030. Defense Technical Information Center
2. Cantwell HR (2009) Operators of air force unmanned aircraft systems: breaking paradigms. Airpower J 23(2):67
3. Dempsey ME (2010) Eyes of the army–US army roadmap for unmanned aircraft systems 2010–2035. US Army UAS Center of Excellence, Ft. Rucker, Alabma, 9

4. Deptula D, Mathewson E (2009) Air force unmanned aerial system (UAS) Flight Plan 2009–2047. Air Force Washington Dc Director Intelligence Surveillance And Reconnaissance
5. Hing JT, Oh PY (2009) Development of an unmanned aerial vehicle piloting system with integrated motion cueing for training and pilot evaluation. In: Unmanned Aircraft Systems. Springer, Netherlands, pp 3–19
6. Menda J, Hing JT et al (2011) Optical brain imaging to enhance UAV operator training, evaluation, and interface development. J Intell Rob Syst 61(1–4):423–443
7. Schreiber BT, Lyon DR, et al (2002) Impact of prior flight experience on learning predator UAV operator skills. Air Force Research Lab Mesa Az Human Effectiveness Directorate
8. The Official Website of the US Air Force. RQ-4 GLOBAL HAWK. http://www.af.mil/information/factsheets/factsheet.asp?id=13225. Accessed 21 June 2013
9. The Official Website of the US Air Force. MQ-9 REAPER. http://www.af.mil/information/factsheets/factsheet.asp?id=6405. Accessed 21 June 2013
10. The Official Website of the US Air Force. MQ-1B PREDATOR. http://www.af.mil/information/factsheets/factsheet.asp?id=122. Accessed 21 June 2013
11. The Official Website of the US Air Force. SCAN EAGLE. http://www.af.mil/information/factsheets/factsheet.asp?id=10468. Accessed 21 June 2013
12. The Official Website of the US Air Force. RQ-11B RAVEN. http://www.af.mil/information/factsheets/factsheet.asp?id=10446. Accessed 21 June 2013
13. The Official Website of the US Air Force. WASP III. http://www.af.mil/information/factsheets/factsheet.asp?id=10469. Accessed 21 June 2013
14. Tirre WC, Hall EM (1998) USAF air vehicle operator training requirements study
15. Under Secretary of Defense for Acquisition, Technology and Logistics (2012) Future unmanned aircraft systems training, operations, and sustainability. Department of Defense
16. Wikipedia, The Free Encyclopedia. File:RQ-11 Raven 1.jpg. http://en.wikipedia.org/wiki/File:RQ-11_Raven_1.jpg. Accessed 21 June 2013
17. Wikipedia, The Free Encyclopedia. File:MQ-1 Predator controls 2007-08-07.jpg. http://en.wikipedia.org/wiki/File:MQ-1_Predator_controls_2007-08-07.jpg. Accessed 21 June 2013
18. Zeng YP, Zhu LL (2010) US Military Unmanned Aircraft System Training. Trainer 4:016
19. Zhu HY, Niu YF et al (2010) State of the art and trends of autonomous control of UAV systems. J Natl Univ Defense Technol 32(3):115–120

Application and Comparison of Imputation Methods for Missing Degradation Data

Ye Fan, Fuqiang Sun and Tongmin Jiang

Abstract A common problem in accelerated degradation testing (ADT) and prognostic and health management (PHM) is the missing of degradation data caused by failure of data transmission or manipulation errors. Facing with such cases, the missing data is usually ignored or even the whole group of data is abandoned. And the loss of valuable information may leads to inaccurate result in the following work. At present, there are various imputation methods have been applied to handling missing data in the field of statistics. These methods estimate the missing values by utilizing the observed data. Unlike most statistical data, degradation data changes over time. But the observed degradation data can still provide valuable information for the estimating. It is therefore reasonable to use these imputation methods to deal with the missing degradation data. The purpose of this paper is to investigate the possibility of using these methods for estimating missing values in degradation data. The missing mechanisms of degradation data are studied at first. Then three of the most widely used imputation methods are researched and used. And comparisons are carried out to show the efficiency of the three methods.

Keywords Degradation data · Missing data · Imputation method · Mean imputation · Regression imputation · Expectation maximization

Y. Fan · T. Jiang
School of Reliability and Systems Engineering, Beihang University,
37 Xueyuan Road, Haidian District, Beijing 100191, China
e-mail: fy@dse.buaa.edu.cn

T. Jiang
e-mail: jtm@buaa.edu.cn

F. Sun (✉)
Science and Technology on Reliability and Environmental Engineering Laboratory,
Beihang University, 37 Xueyuan Road, Haidian District, Beijing 100191, China
e-mail: sunfuqiang@buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_137

1 Introduction

In life prediction or failure prediction of accelerated degradation testing (ADT) and prognostic and health management (PHM), traditional data processing methods, like time series algorithm, need the complete degradation data to be the input. Missing values of degradation data are often encountered due to the failure of data transmission or manipulation errors. In such cases, the traditional data processing methods can do nothing with it. Then the whole group of data with missing values is always abandoned. Obviously, it is a waste of data resource. Some data processing methods can use the incomplete data as the input, but the loss of valuable information may leads to inaccurate result. Then the inaccurate result will have an effect on the following work like life prediction or failure prediction.

Various methods are proposed to handling the missing data problem in the field of statistics [1, 3, 4]. The most popular technique is imputation, which estimates the missing values by utilizing the observed data to make the data complete [2, 5]. And imputation methods also strive to make the data as real as possible.

In the case of missing degradation data, imputation methods can be applied. A reasonable imputation method can make better use of the data, and may even influence the conclusion. So it is a key step to deal with missing degradation data.

In this chapter, the missing data mechanism of degradation data is discussed firstly. The application of three popular imputation methods to degradation data are researched secondly. In order to compare the efficiency of the three methods, statistical analysis is conducted thirdly. At last, we conclude by discussing the findings.

2 Missing Data Mechanisms

The methods of handling missing data are directly related to the mechanisms that caused the incompleteness.

In general, three types of missing data mechanisms exist [4]. (1) Missing completely at random (MCAR): The missing data is unrelated to the values of any other data, whether observed or missing. (2) Missing at random (MAR): The missing data is related to the values of the observed data, and not to the missing ones. (3) Not missing at random (NMAR): The missing data depends on both on the values of the observed data and the values of missing data.

Because degradation data has a certain degradation process, data is related to each other. And the missing degradation values depend on the observed part, not the missing part. So the missing data mechanism of degradation data is MAR.

3 Imputation Methods

At present, there are various imputation methods have been applied to handling missing data in the field of statistics. Mean imputation (MI), regression Imputation (RI) and expectation maximization (EM) are three of the most widely used imputation methods [1, 2]. And they are capable to deal with the missing data of MAR.

3.1 Mean Imputation

Mean imputation replaces the missing values with the mean of the observed values. In the case of normal distribution, the sample mean provides an optimal estimate of the most probable value. Because degradation data changes with a certain path, the mean of observed values cannot be used directly and data transformation is necessary. If degradation data is linear, first difference transformation should be done firstly. If degradation data is nonlinear, then data transformation depends on the actual conditions. The imputation value of MI is as follows

$$Y'_{MI} = \bar{Y}'_{obs} \tag{1}$$

where Y'_{obs} denotes the data after transformation $\bar{Y}'_{obs}$ denotes the mean of Y'_{obs}. Then complete degradation data will be obtained after Y'_{MI} is transformed back.

Although one may impute all missing values, the variance of degradation data will shrink, because the mean imputation value will contribute nothing to the variance. So in this paper a random term is added to the imputation value to keep the variability of data as the same of original as possible. Then the modified imputation value is

$$Y^*_{MI} = Y_{MI} + \varepsilon \tag{2}$$

where $\varepsilon \sim N(0, \sigma^2)$, σ^2 is the variance of the observed data, Y_{MI} denotes the traditional imputation value. The process of MI is shown in Fig. 1.1.

3.2 Regression Imputation

Regression Imputation utilizes the regression relationship between the observed degradation data and variables $x_k (k = 1, 2, \ldots, m)$. Regression model is established by the application of multiple regression. The missing values are estimated through the known variables and the regression model. If the relationship between the degradation data and variables is linear, the ith imputation value of RI is

$$y_{RI\,i} = \beta_0 + \sum_{k=1}^{m} \beta_k x_{k\,i} + \varepsilon_i \tag{3}$$

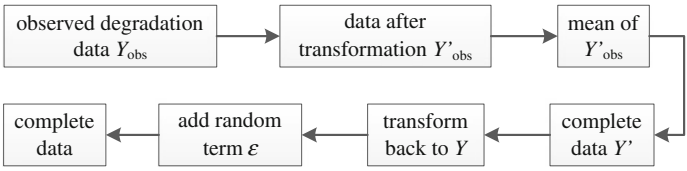


Fig. 1 Flow chart of MI

Where β denotes the regression coefficients, ε is the error term, $\varepsilon \sim N(0, \sigma^2)$, σ^2 is the variance of the regression residuals from the observed degradation data.

If the relationship between the degradation data and variables is nonlinear, a model can still be found, and the ith imputation value is

$$y_{RIi} = \sum_{k=1}^{k} f(x_{ki}) + \varepsilon_i \tag{4}$$

where $f(x_k)$ denotes the model of degradation data and variable x_k. The process of MI is shown in Fig. 1.2.

3.3 Expectation Maximization

The EM algorithm is an iterative algorithm that finds the parameters that maximize the logarithmic likelihood when there are missing data. Each iteration consists of an E-step (expectation step) and M-step (maximization step). The E-step calculates the conditional expectation of the complete data log likelihood, given the observed data and the current parameter estimates. Suppose $^{(t)}$ is the current estimate of the parameter θ. Then

$$Q\left(\theta|\theta^{(t)}\right) = \int l(\theta|y) f\left(Y_{mis}|Y_{obs}, \theta = \theta^{(t)}\right) dY_{mis} \tag{5}$$

where $l(\theta|y)$ is the complete data log likelihood. Given complete data log likelihood, the M-step finds the parameter estimates to maximize the complete data log likelihood from E-step.

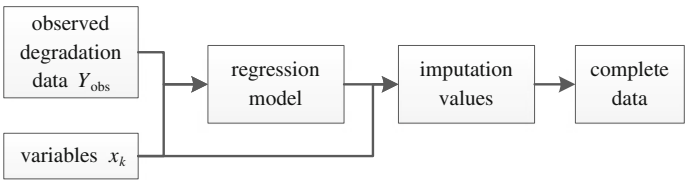


Fig. 2 Flow chart of RI

$$Q\left(\theta^{(t+1)}|\theta^{(t)}\right) \geq Q\left(\theta|\theta^{(t)}\right) \text{ for all } \theta \tag{6}$$

The E-step and the M-step are iterated until the iteration converges. Because degradation data do not obey any distribution, data transformation is necessary before EM is applied. Then Y_{obs} is transformed to Y'_{obs} which obeys a certain distribution. Degradation data is always a large amount of data. According to the law of large numbers, Y'_{obs} is considered to obey normal distribution. Then normality test need to be done after data transformation to ensure Y'_{obs} is normally distributed. If normality test is unqualified, data transformation should be redone. If normality test is qualified, EM can be used.

Suppose $y'_i \sim i.i.d\ N(\mu, \sigma^2)$, for $i = 1,2,\ldots, r$, y'_i is observed, for $i = r+1, r+2, \ldots, n$, y'_i is missing. Given Y'_{obs} and $\theta = (\mu, \sigma^2)$, the expectation of each y'_i is μ, and $l(\theta|y)$ is a function of $\Sigma y'_i$ and $\Sigma y'^2_i$ which are sufficient statistic. Suppose $\theta^{(t)} = (\mu^{(t)}, \sigma^{(t)})$ is the current estimate, then the E-step is

$$E\left(\sum_{i=1}^{n} y'_i|\theta^{(t)}, Y'_{obs}\right) = \sum_{i=1}^{r} y'_i + (n-r)\mu^{(t)} \tag{7}$$

$$E\left(\sum_{i=1}^{n} y'^2_i|\theta^{(t)}, Y'_{obs}\right) = \sum_{i=1}^{r} y'^2_i + (n-r)\left[(\mu^{(t)})^2 + (\sigma^{(t)})^2\right] \tag{8}$$

And M-step is

$$\mu^{(t+1)} = E\left(\sum_{i=1}^{n} y'_i|\theta^{(t)}, Y'_{obs}\right) \Big/ n \tag{9}$$

$$\left(\sigma^{(t+1)}\right)^2 = E\left(\sum_{i=1}^{n} y'^2_i|\theta^{(t)}, Y'_{obs}\right) \Big/ n - \left(\mu^{(t+1)}\right)^2 \tag{10}$$

when iteration converges, the estimate of $\theta = (\mu, \sigma^2)$ is figured out. The missing values will be imputed according to the μ and σ^2. And the complete data Y will be obtained after Y' is transformed back. The process of EM is shown in Fig. 1.3.

4 Comparison of Imputation Methods

Comparisons are carried out under different percentages of missing data to show the efficiency of the proposed methods. MI, RI and EM are applied to missing degradation data which is simulated, when the missing rate are 10, 20, 30, 40, 50 and 60 % (as shown in Fig. 1.4). The evaluation criterion includes absolute error,

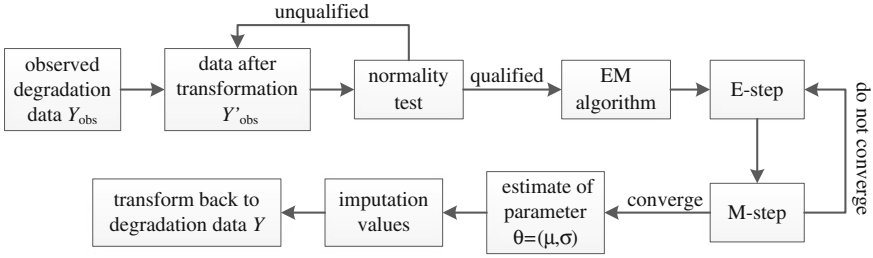


Fig. 3 Flow chart of EM

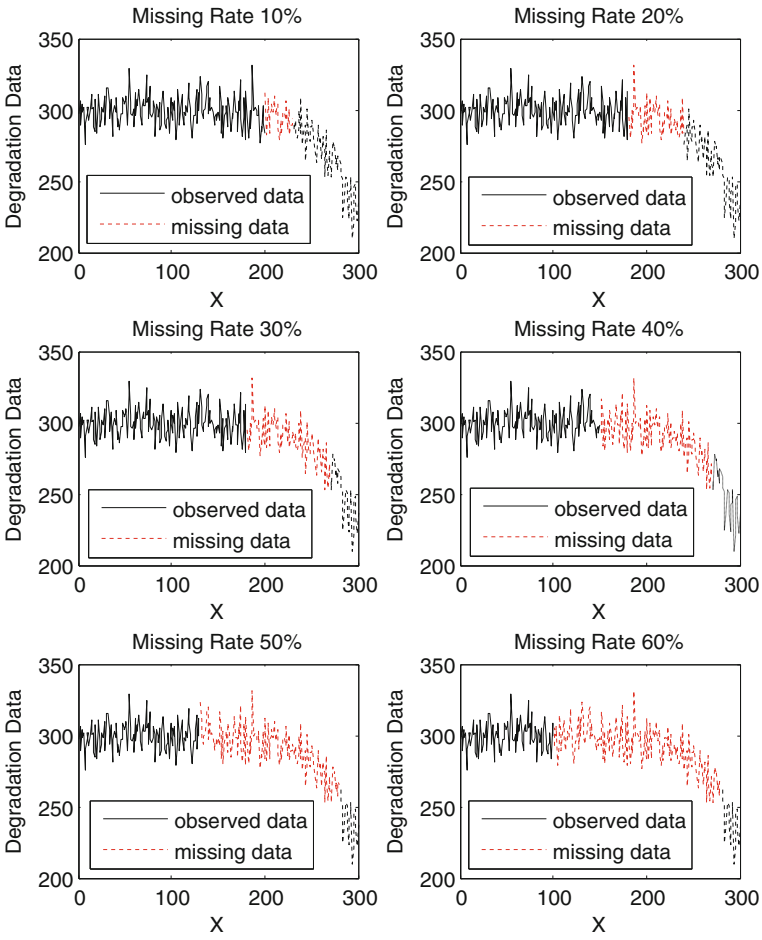


Fig. 4 Degradation data with different missing rate

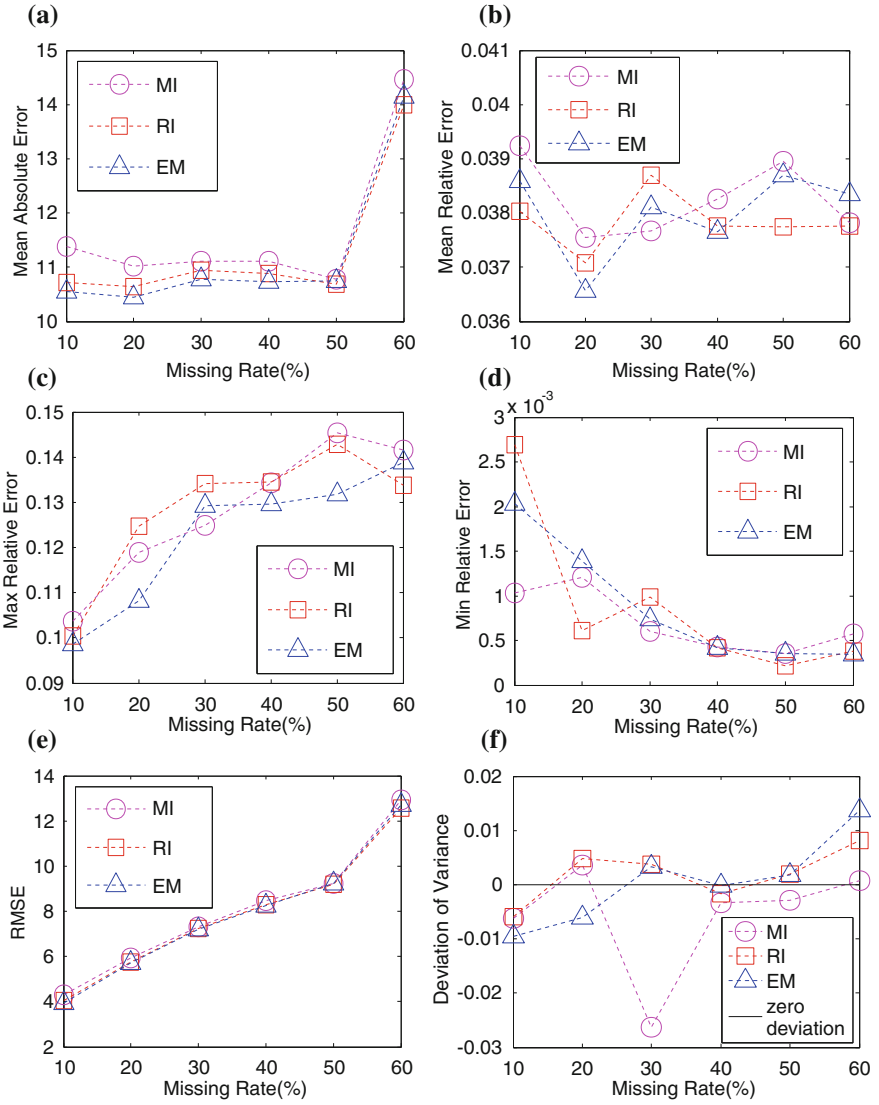


Fig. 5 Comparison of MI, RI and EM

relative error, max relative error, min relative error, root mean square error (RMSE) and deviation of variance. The results of comparisons are shown in Fig. 1.5.

From Fig. 1.5a–e we can conclude that MI is not best method in most cases; when missing rate is under 50 %, EM is the most efficient method; when missing rate is 50 %, EM and RI are nearly the same good; when missing rate is up to 60 %, the efficiency of EM decreases, and RI becomes the best.

Figure 1.5f shows the deviation of variance of imputation data with respect to the variance of original data. The variance of data reflects the degree of dispersion of data. The variance of imputation data obtained by a good method is supposed to be as close to the original as possible [5]. Although a random term is added to the imputation value of MI, the variance still shrinks. The variance of imputation data obtained by RI is closest to the original variance, and it is stable relatively. The EM's deviation of variance is slightly bigger than RI's.

5 Conclusion

In this chapter we investigate the possibility of using imputation methods for estimating missing values in degradation data. MI, RI and EM are researched and used, and the results show that they all have the ability to handle the missing degradation data. MI is easy to be carried out, but its efficiency is not high. RI does not need data transformation, and its efficiency is better than the other two when the percentage of missing data is high. EM is difficult to be carried out relatively, but its efficiency is always the best of the three when the percentage of missing data is low.

References

1. Barzi F, Woodward M (2004) Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. Am J Epidemiol 160:34–45
2. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Roy Statist Soc B 39:1–38
3. Little RJA, Rubin DB (1983) Incomplete data. Encycl Statist Sci 4:46–53
4. Little RJA, Rubin DB (2002) Statistical analysis with missing data, 2nd edn. John Wily and Sons, New York
5. Rubin DB (1987) Multiple imputation for nonresponse in surveys. John Wiley and Sons, New York

A Simulation Research on Test Point Selection for Analog Electronic Systems on Diagnosis and Prognosis

Jiaming Liu, Shunong Zhang and Shuang Xie

Abstract The selection of test points is one of the key steps for electronic systems on prognostic and health management (PHM). Parameters of reasonable test points can characterize the fault features of electronic systems and provide effective input information for diagnosis and prognosis. In this chapter, an improved method of test point selection for analog electronic systems on diagnosis and prognosis is proposed and a case study is presented. The proposed method includes several steps: (1) fault risk analysis: determine susceptibility areas for the electronic system based on environmental stresses analysis and life analysis; (2) functional simulation analysis: based on the circuit schematic diagram, establish the functional simulation model for the system and list all accessible nodes, then identify the relationship between nodes and faults by using correlation models; (3) fault simulation analysis: Combine with the previous analysis of failure modes and mechanisms and set corresponding faults in the functional simulation software; (4) Comprehensive evaluation and selection: evaluate fault simulation data of each test point, and select appropriate test points for the system. The presented method has the following features: considering the specific product condition and failure mechanisms, focusing on high-risk mechanisms, and basing on fault simulation to revise testability model to make it closer to the real fault situations.

Keywords Test point selection · Diagnosis · Prognostic · Simulation · Analog electronic systems

J. Liu (✉) · S. Zhang · S. Xie
National Laboratory for Reliability and Environmental Engineering, School of Reliability and Systems Engineering, Beihang University, Beijing, China
e-mail: liujiaming@dse.buaa.edu.cn

S. Zhang
e-mail: zsn@buaa.edu.cn

S. Xie
e-mail: yeerxs@dse.buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_138

1 Introduction

For electronic systems, test points are those points extracting system information. Parameters of reasonable test points can characterize the fault or fault features for electronic systems and provide effective input information for diagnosis and prognosis. Test point selection for an electronic system should be conducted in the stage of design, during which circuit functional simulations and fault simulations are effective methods to verify if the circuit functions are correct and to find the errors of design of the system and then to revise them. Generally, the software of a circuit functional simulation includes the function of circuit fault simulation and the circuit fault simulations is a core technology in a circuit functional simulation. The purpose of circuit fault simulations is to find possible fault modes and mechanisms of the electronic system under certain operating conditions. Circuit functional simulation software in common use includes Multisim, Aitum/Protel, ORCAD and PSPICE, etc.

Since 1980s, model-based testability analysis technology has been emerging, which can evaluate the testability of an electronic system and optimize test points by abstracting the electronic systems (including components and testing) to logical models. Here the test points mean the abstract nodes providing testability information, and the testing is not limited to the testing of electrical parameters, but also that of temperature, humidity and other physical parameters [1]. Currently, the representative models and their software tools include the correlation model and its testability design software TEAMS from DSI Company, the signal flow model and its system testability and maintenance platform (STAMP) from ARINC Company and the multi-signal flow model from the University of Connecticut [2].

However, the method of circuit fault simulations and the method of model-based testability analysis have the following deficiencies: (1) it mainly focuses on fault diagnosis situations; (2) generally, it regards all faults as hard faults; (3) the testing result for a fault is only pass or not pass two kinds.

In recent years, with the development of prognosis technology, the test point selection for electronic systems meets new requirements. Test points not only need to provide information for diagnosis, but also need to provide information for prognosis.

Currently, the popular prognostic methods for electronic systems can be divided into two categories: (1) based on Physics Of Failure (POF) models; (2) Based on Data Driven (DD) methods. For the POF based method, generally the damage accumulation caused by a variety of failure mechanisms on a product needs to be calculated by real-time monitoring and getting the working stresses (e.g. voltage, power, etc.) and environmental stresses (e.g. temperature, vibration, etc.) from test points of the product for the purpose of further prognosing the Time To Failure (TTF) of the product. For the DD methods, some characteristic parameters (typically, e.g. leakage current, output voltage, operating current and so on) reflecting some changes in the performance levels or health/fault status for an electronic system are needed to prognose the TTF or remaining useful life for the system by some internal relations between the failure modes and the parameter changing or

with some parametric or nonparametric algorithms. So it requires the test points selected to be able to characterize the fault precursors of the system and to be sensitive enough when the performance degradation of the system or its components occurs.

This chapter makes an attempt to provide an improved method based on the existing model-based method of test point selection to meet the requirements of fault diagnosis and also prognosis for analog electronic systems.

2 Steps of Test Point Selection

The process of test point selection is shown in Fig. 1, which mainly includes the following four steps:

1. Fault risk analysis: determine susceptibility areas for the electronic system based on environmental stresses analysis and life analysis.

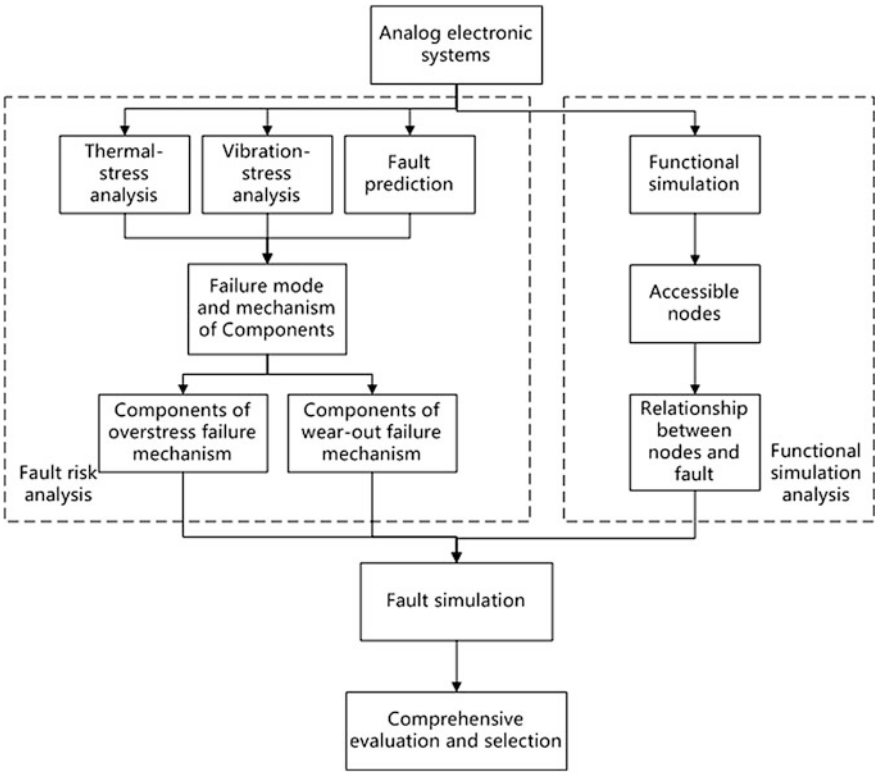


Fig. 1 Processes of test point selection

2. Functional simulation analysis: based on the circuit schematic diagram, establish the functional simulation model for the system and list all accessible nodes, then identify the relationship between nodes and faults by using correlation models.
3. Fault simulation analysis: combine with the previous analysis of failure modes and mechanisms and set corresponding faults in the functional simulation software.
4. Comprehensive evaluation and selection: evaluate fault simulation data of each test point, and select appropriate test points for the system.

Each step above is illustrated below through a case study with a local analog circuit module of a board system.

2.1 Fault Risk Analysis

Here, the voltage conversion module in a board system, whose circuit schematic diagram is shown in Fig. 2, is as an example to illustrate the steps of test point selection above. The input signals in this module are +10, +15 and −15 V, while the output is −13.4 V. The module is composed of the following nine components: resistors R307, R309, R311 and R312; capacitor C111 and C203; transistors Q30 and Q300; integrated power amplifier N318.

Figure 3 shows the PCB diagram of the whole board system. All components in the voltage conversion module in Fig. 2 are located in the area of rectangle frame in red near the right edge.

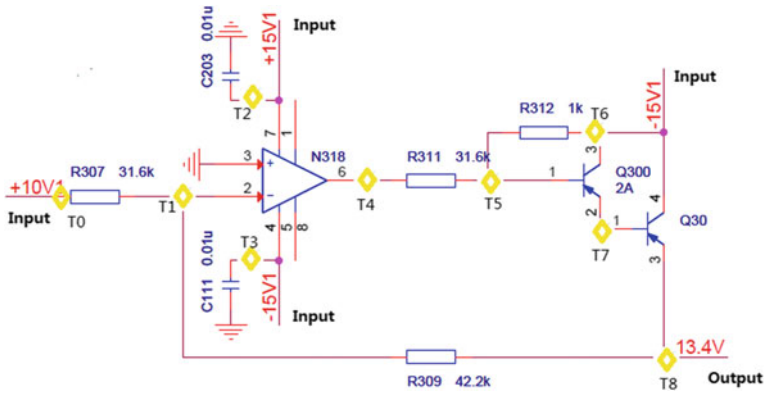


Fig. 2 Circuit schematic diagram of the voltage conversion module

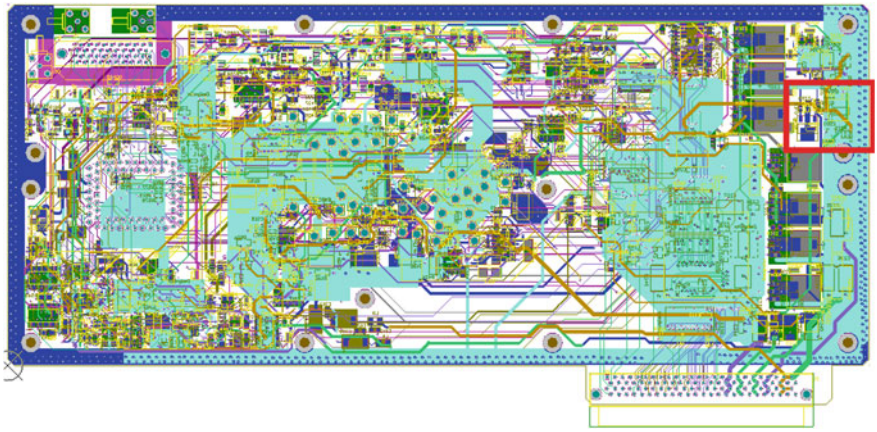


Fig. 3 The PCB diagram of the whole board system

2.1.1 Identify Life Cycle Loads

The circuit board is one part of an equipment system where the environmental temperature usually is about 50 °C in operation and 25 °C in closing down. The rates of change from 25 to 50 °C and from 50 to 25 °C are both 5 °C/Min. The equipment is operated once a day, and the lasting time is 180 min each time. Other environmental loads can be ignored.

2.1.2 Classification of Failure Mechanisms

Generally, the failure mechanisms are divided into two types: wearout and overstress. Overstress failures involve a failure that arises as a result of a single load (stress) condition, while wearout failures on the other hand involve a failure that arises as a result of cumulative load (stress) conditions. For the overstress failure mechanism, a stress analysis is needed to determine if a failure is precipitated under the given environmental and operating conditions, and enough margin design needs to be considered. For the wearout mechanisms, the TTF needs to be calculated under the given environmental and operating conditions.

2.1.3 Stress Analysis

Flotherm software can be used to analyze the thermal stress on the circuit board. The result of thermal stress analysis for the circuit board is shown in Fig. 4. It can be seen that the high temperature area concentrates in the upper right part.

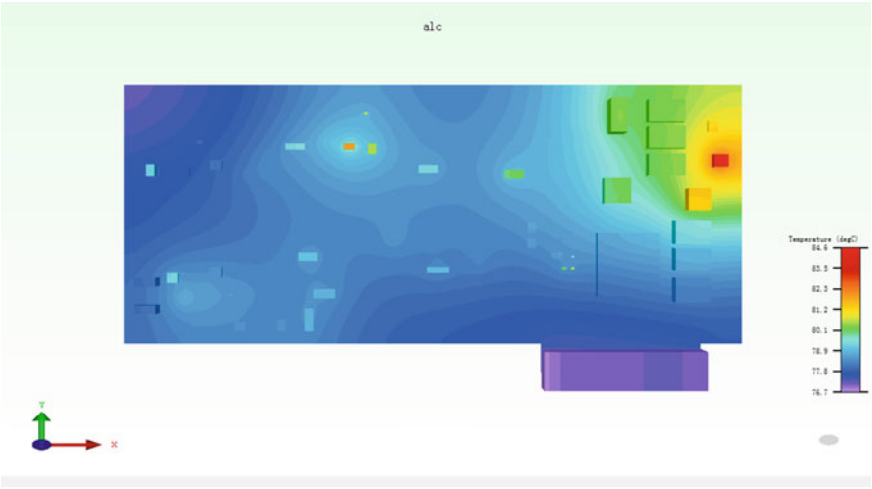


Fig. 4 Results of thermal stress analysis

2.1.4 Prediction for TTFs and Evaluation for Risk Levels

For each component, the potential failure modes, failure causes and failure mechanisms need to be identified, and the related failure models need to be found. Then the TTF for each component can be calculated by the related model. In cases where no failure models are available, the evaluation is based on past experience, manufacturer data or handbooks. The risk level for each component also needs to be considered, which get from occurrence levels and severity levels. Table 1 shows the TTF and the risk level for each component of the voltage conversion module. The high-risk components, which are N318, R307 and R309 in the voltage conversion module for the example, need to be focused on. The calculation can be conducted by CalcePWA and CalceFAST software which are developed by the CALCE of the University of Maryland.

2.2 Preliminary Selection of Test Points

For the voltage conversion module, a functional simulation model can be established by using Multisim software according to the circuit schematic diagram. Nine accessible nodes which are T0 ~ T8 in yellow and diamond shapes in Fig. 2 are available. Assuming that the input circuit is normal, then the five nodes T1, T4, T5, T7 and T8 can be as a test point respectively for further analysis because the nodes T0, T2, T3 and T6 are on the input positions.

Figure 5 shows the testability model of the voltage conversion module, which is established by a modeling software TMAS (Testability Modeling and Analysis System) developed by the RSE in Beihang University. The round shapes in yellow

Table 1 TTF and risk level for each component of the voltage conversion module

Components	Potential failure mode	Potential failure cause	Potential failure mechanism	Mechanism type	Failure Model	TTF	Occurrence level	Severity level	Risk level
R307	Solder open	Thermal cycling	Fatigue	Wear-out	First Order Thermal Fatigue Model	29.3 year	Occasional	High	High
R309	Solder open	Thermal cycling	Fatigue	Wear-out	First Order Thermal Fatigue Model	29.7 year	Occasional	High	High
R311	Solder open	Thermal cycling	Fatigue	Wear-out	First Order Thermal Fatigue Model	>30 year	Remote	High	Moderate
R312	Solder open	Thermal cycling	Fatigue	Wear-out	First Order Thermal Fatigue Model	>30 year	Remote	High	Moderate
C111	Solder open	Thermal cycling	Fatigue	Wear-out	First Order Thermal Fatigue Model	29.4 year	Occasional	Moderate/ significant	Moderate
C204	Solder open	Thermal cycling	Fatigue	Wear-out	First Order Thermal Fatigue Model	29.4 year	Occasional	Moderate/ significant	Moderate
Q30	Solder open	Thermal cycling	Fatigue	Wear-out	First Order Thermal Fatigue Model	>30 year	Remote	High	Moderate
Q300	Solder open	Thermal cycling	Fatigue	Wear-out	First Order Thermal Fatigue Model	>30 year	Remote	High	Moderate
N318	Short circuit	Thermal cycling	Die fatigue crack	Wear-out	Westergaard Bolger Model	3.36 year	Reasonably Probable	High	High

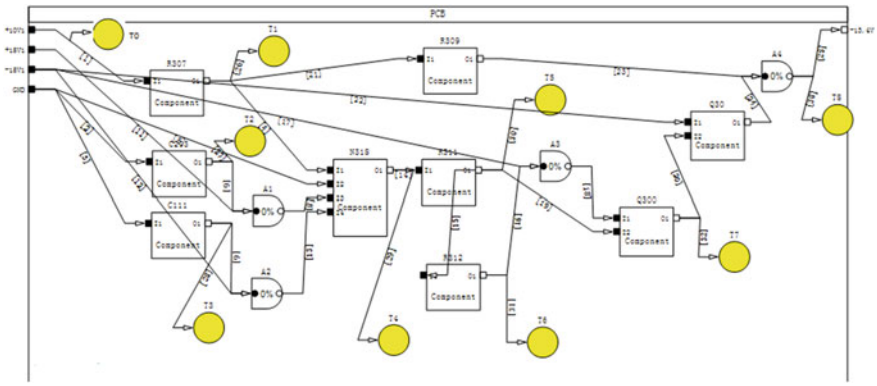


Fig. 5 The testability model of the voltage conversion module

in Fig. 5 are corresponding to the diamond shapes in yellow in Fig. 2. The model shows the flow of signals and the relationships between nodes and faults. Assuming that all components may occur faults, then a test matrix can be obtained for the voltage conversion module, as shown in Table 2. The values in the table indicate the test results from test points for fault sources. "0" shows that the test point cannot detect out faults, versus "1" indicates that the test point can detect out faults.

2.3 Fault Simulation

Through the failure mechanism analysis in Sect. 2.1, three high risk components are indicated. Then simulations in different fault conditions for the N318, R307 and R309 are needed to verify the relationship between test points and faults in Table 2, and evaluate the test point performances.

Table 2 A test matrix for the voltage conversion module

Fault sources	T1	T4	T5	T7	T8
R312	0	0	0	1	1
R307	1	1	1	1	1
C111	0	1	1	1	1
N318	0	1	1	1	1
R309	0	0	0	0	1
C203	0	1	1	1	1
R311	0	0	1	1	1
Q30	0	0	0	0	1
Q300	0	0	0	1	1

For R307 or R309, the fault gradual process can be shown as the increased resistance, and the health statuses can be defined as three levels:

(a) Normal status: a resistance s in 5 % of nominal value;
(b) Soft fault status: a resistance is in 5–30 % of nominal value;
(c) Hard fault status: a resistance is more than 30 % of nominal value;

Their failure mechanisms are the solder joint fatigues caused by temperature cycling. In order to simulate the fault gradual process, 0, 10, 20, 30, 40 and 50 % exceeded the nominal resistance values are set. The expect health status of the components is from normal status to soft fault status, and then to hard fault status. Figures 6 and 7 show the values of the parameters (U1, U4, U5, U7 and U8) of the test point T1, T4, T5, T7 and T8 for R307 and R309 respectively. It is can be seen that the test point T1, T5, T7, and T8 can detect the fault; however, T1 is the most sensitive test point. So, the test matrix should be revised as shown in Table 3.

2.4 Comprehensive Evaluation

The main purpose of the test point selection is to focus on high risk of fault (N318, R307, and R309). For N318, it is hard to test the gradual fault by the parameter of voltage, however, if its failure causes the component to short-circuit, only the test point T4 can detect the fault, so choose T4 as one of the final test points. For R307 and R309 fault, the test matrix does not give enough information to select test points. Based on the previous analysis, T1 is the most sensitive test point. However, for R307 and R309 fault, if only selecting T1 as test point, we can't identify

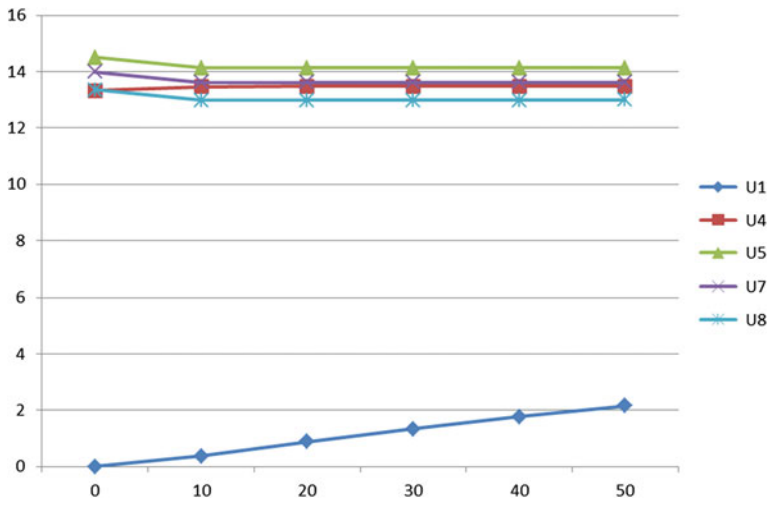


Fig. 6 Test points response for R307 faults

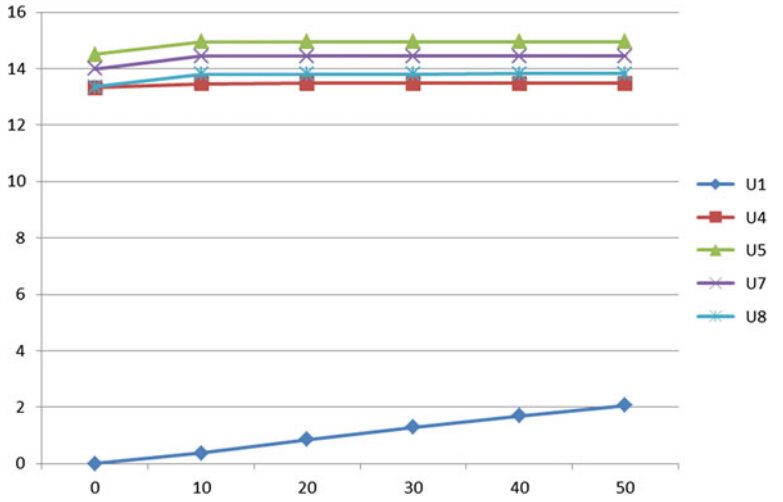


Fig. 7 Test points response for R309 faults

Table 3 The revised test matrix

Fault sources	T1	T4	T5	T7	T8
R312	0	0	0	1	1
R307	1	0	1	1	1
C111	0	1	1	1	1
N318	0	1	0	0	0
R309	1	0	1	1	1
C203	0	1	1	1	1
R311	0	0	1	1	1
Q30	0	0	0	0	1
Q300	0	0	0	1	1

distinguish R307 fault and R309 fault when parameter of T1 changed. So a test point is needed to add to achieve fault isolation. When R307 or R309 fault occurs, T5, T7 and T8's response is different. Considering T8 is the output point of the module, we can choose T8 as test point for isolation. So, T1, T4 and T8 are selected as final test points for this voltage conversion module for the purpose of detecting or predicting the high risk faults.

Discussion: The test matrix is revised based on the fault simulation results. Comparing with existing model-based testability modeling methods, this method to select test points can effectively solve the problem of inaccurate modeling. However, it also has a problem that huge simulations are needed for the system, so it is more suitable for use in each module.

3 Conclusion

This chapter presents an improved method for test point selection on fault diagnosis and prognosis on electronic systems based on the model-based method for test point selection. The method has the followin features to select test points: (1) it is considered the specific product condition and failure mechanisms; (2) it focuses on high-risk mechanisms; (3) it is based on fault simulation to revise testability model to make it closer to the real fault situations.

References

1. Jiang RH, Long B, Wang HJ (2007) Test selection based on improved binary particle swarm optimization. doi:10.1109/ICIEA.2007.4318675
2. Zhang Y, Qiu J, Liu GH (2011) Comparison and prospect of testability models. J Test Meas Technol 25:504–514 (in Chinese)

The Human Dimension of Asset Management

David van Deventer

Abstract Utilities around the world use Asset Management Plans (AMPs) in some form, as a key tool to direct, operate and manage their assets over time. This paper will examine the degree to which human asset planning and management is aligned with, or incorporated in AMP's, with specific reference to practices in New Zealand. In particular, it will examine and report on: *1.* The nature of New Zealand and its utilities as a useful model for analysis, being a compact First World country with a diverse population and steadily changing demographic. *2.* WEL Networks as an integrated distribution network company that employs all the human assets required to plan, operate and maintain its assets. *3.* The need for incorporating human asset planning into the AMP. *4.* The components and variables of human asset planning and plans. *5.* The role of institutional knowledge and intellectual property in sustaining performance. *6.* Some methodologies developed for analysing dimensions of human asset planning, such as: a. demand forecasting b. demographic status and analysis of their impacts c. competence assessment and d. identification of development potential. *7.* Some methodologies for optimising human assets: a. Creating an employer brand and employee value proposition b. Longer-term development strategies c. Employee engagement and motivation for performance. The paper will conclude with some key success factors and practices for adding value to Asset Management Plans by incorporating Human Asset Planning as an integrated component of Asset Management.

Keywords Diversity · Demographics · Human asset planning · Competence · Development potential · Employer brand · Employee engagement · Institutional knowledge (IP)

D. van Deventer (✉)
WEL Networks Limited, Hamilton, Waikato, New Zealand
e-mail: David.vandeventer@wel.co.nz

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_139

1 Introduction

Utilities around the world use Asset Management Plans (AMP's) in some form, as a key tool to plan investment, operate and manage their assets. This paper examines the degree to which human asset planning and management can be aligned with AMP's, with specific reference to practices at WEL Networks Limited (WEL) in New Zealand.

2 Background

New Zealand is a compact First World country with a small (4 million) but diverse population and steadily changing demographics. WEL is an electricity distribution network company that employs all the human assets required to plan, operate and maintain its assets to deliver a reliable and cost effective power supply safely to just over 85,000 homes and businesses.

Safety, reliability, quality and value for money are key requirements and aligning human asset planning with the AMP can make a significant contribution to achieve these requirements.

3 Work Done and Methodology

3.1 Human Asset Demand Forecasting

Forecasting human resource demand from an AMP is relatively easy. The required Knowledge, Skills and Behaviours can be deduced at high level from the deliverables defined in the AMP. The requirements or criteria against each of these are normally defined in a job description as Education (Knowledge), Skills (Experience and training) and Competencies or Personal Attributes (Behaviours).

The quantity required and timing can be determined from the implementation plans for the AMP.

The approach to forecasting typically differs between companies, depending on whether they have adopted an out-sourced or in-sourced resourcing model. In the out-sourced model the contractor would base their resource forecasting on meeting the contractual delivery requirements for the duration of the contract, whilst remaining competitive in a competitive tender process.

In the in-sourced model the planning horizon and hence forecasting is not based on the finite term of a contract but rather on efficiency whilst ensuring continuity of resourcing and performance.

3.2 Human Asset Supply Forecasting

Due to a significant number of variables, supply can be forecast but not confidently predicted. A significant factor influencing forecasting ability and accuracy is whether the human assets are an integral part of the company or contracted from another party.

An equally important consideration is the role of intellectual property and corporate memory. This is addressed later in the paper.

3.2.1 Out-Sourced Human Assets

This is a model used in many industries and businesses. The deliverables are defined in a contract and the Contractor has to meet the combination of quantity, quality, time and cost standards specified. Minimum educational qualifications, experience or similar requirements may be specified in some instances but the Contractor determines the human asset supply required to meet the Contractual requirements, whilst making an acceptable commercial return.

3.2.2 In-Sourced Human Assets

WEL Networks adopted this model in 2008 when resources for Fault Response, Maintenance and Capital Replacement work were in-sourced. Major New Works remain primarily outsourced due to the variability of the work quantum and timing, but the company employs the resources for the Project Management of both replacement and new capital works.

A significant component of supply forecasting is to measure and track human asset data. This includes:

1. Base data on qualifications, registrations, certifications and endorsements.
2. Work and experience history.
3. Median age and age spread, to predict loss due to retirement, but also the potential impact of aging on physical ability, especially in manual handling and climbing.
4. Staff turnover by role type, reason, tenure and seniority level. Data gathered in exit interviews provide further context.
5. Development potential.

4 Determination of Development Potential

This is an important component of human asset supply forecasting for a number of reasons:

1. Development investment can be targeted to meet both business needs and personal needs.
2. A developmental pathway allows career progression and contributes positively to both the employer brand and staff engagement and retention
3. Intellectual property and corporate memory grows with the individual and is retained within the business
4. It is possible to forecast when the developing capability will be ready for deployment
5. It supports succession planning for key roles, including the planned transfer of intellectual property and corporate memory

4.1 Identifying Development Potential

This could evolve into a very complex process where the investment would not justify the results achieved.

However, at WEL we used the research on identifying High Potential Individuals (HIPO's), done by the Corporate Leadership Council (CLC) in the United States [1]. This forms the basis for our assessment model used to quantify and prioritise development potential.

The CLC identified three indicators of development potential; Ability, Engagement and Aspiration.

WEL took this basic concept and added the following additional factors; Qualifications, Tenure, Age and Leadership ability.

4.1.1 Definitions

1. Ability—The innate intellectual and cognitive ability to learn or do new things
2. Aspiration—A conscious desire to advance for material and/or prestige benefits
3. Engagement—An emotional and rational commitment that remaining with the company is in their own best self interest
4. Qualification—The assessment level of a qualification as determined by the New Zealand Qualification Authority (NZQA)
5. Leadership—The ability and desire to direct the efforts and maximise the performance of people

6. Tenure—Years service, as a trigger for starting appropriate development investment
7. Age—Years, as a determinant of expected remaining period of tenure to benefit from the development investment

4.1.2 Weighting of Individual Factors

Because of the diversity of factors in the model, we found it very important to get the relative weightings correct so that the score allocated to each factor ensured that the end result supports valid decision making. Weighting was determined by a panel of senior managers, using the following methodology:

1. The panel selected a factor rated as most important, in our model it was Ability. This was allocated a score of 10
2. The panel then selected the next most important factor. A score, relative to the 10 awarded to Ability was allocated to the factor. If it was of equal importance, it could also be scored 10. A score of 8 would indicate the degree of separation between the factors.
3. The same methodology was applied to all the remaining factors and typically factors such as Tenure or Age were be allocated a relatively low score.
4. All the scores were then totalled and a percentage was calculated for each factor.
5. This percentage was entered into the model, which was then ready for evaluating individuals. An example of the full model is illustrated below (Table 1).

5 Methodologies for Optimising Human Assets

As a relatively compact company with 250 staff in total WEL has to, and is also able to be flexible and responsive to opportunities for optimising human asset acquisition and development.

5.1 Human Asset Acquisition

The development of a strong employer brand and employee value proposition (EVP) is invaluable in differentiating the company in a competitive employment market.

The employer brand at WEL is built around a number of basic messages and strategies:

Table 1 HIPO assessment matrix

Emp	Start Date	Service	DOB	Age	Qual	Age	S've	Ability	Engage	Aspire	Ldr's	Total
A	01/01/00	13	1/01/54	59	2	1	2	4	4	2	3	73 %
B	01/01/08	5	1/01/79	34	1	2	2	5	4	4	4	86 %
				Weight	4	2	2	10	8	7	10	
					9 %	5 %	5 %	23 %	19 %	16 %	23 %	100 %

1. The ownership model, being owned by a Trust on behalf of the wider Waikato community, positions the company as a respected community asset
2. Attention is paid to the presentation of staff, equipment and vehicles with regard to quality, appearance and consistent branding. This is particularly important for field staff who work in the public arena every day.
3. The employee value proposition is based on combining all aspects of the employment relationship into a total value proposition. This includes:
4. Industry benchmarked remuneration and reward systems
5. Supportive policies that extend to the family where relevant
6. A strong commitment to the five core values of the company (ABCDE); Agile, Build the business, Care for our staff, our customers and our assets, Do the right thing and Every day home safe. This is underpinned by our strap line; Best in Service, Best in Safety
7. Quality focussed recruitment practices that leave a positive impression with all applicants, regardless of whether they gain employment or not
8. Quality on-boarding and induction processes that support and reinforce the EVP

5.2 Human Asset Development

In addition to the traditional training activities that focus on operational knowledge and skills development, WEL has a number of different strategies over the past five years that are starting to deliver good results:

1. Increasing the number of trainees, as a percentage of the staff complement, to 15 % of field staff. This has had a number of positive results:

    a. Trainees are a cost effective resource where multiple manning is required for safety or regulatory purposes. This also gives them exposure to real time training and experience
    b. Traineeships are relatively scarce because of the upfront investment required, so quality candidates with significant upside potential can be attracted to the company
    c. Good Trainees quickly become productive, yet are a cost effective resource for delivering work programmes
    d. Experienced staff are encouraged to share their knowledge and experience with the Trainees

2. Creating a “pipeline” for talent growth from the bottom up by:

a. Seconding staff to higher level roles and projects, sometimes across business functions. This provides an opportunity to demonstrate their potential without committing to a permanent promotion
b. Employing a selected number of graduates who are first rotated through various functions before being assigned to roles on the basis of identified succession requirements
c. Identifying vulnerabilities in terms of long term contractors "owning" important company IP and putting individuals, identified for development, in place to identify and bring such IP in-house

6 Conclusion

Getting a utility to go beyond the tired old phrase: "Our people are our biggest asset", to applying the same degree of diligence and rigour to managing human assets, as it does to hard assets, has added and is continuing to add value. We believe our human assets know that:

1. The company has a vision and plan in which they are a significant and essential component.
2. They can identify with and share in the employee value proposition.
3. Talent management is a tool used to achieve an optimal alignment between business, operational and personal objectives.
4. They have a direct role in achieving excellence in safety, reliability, quality and affordability.

Reference

1. Corporate Leadership Council (2005) High-potential employee database CIO executive board research

A Simulation Research on Gradual Faults in Analog Circuits for PHM

Shuang Xie, Shunong Zhang and Jiaming Liu

Abstract Prognostic and Health Management (PHM) is a key technology for condition-based maintenance and autonomic support, and electronic systems are one of the main application areas in PHM. An electronic system generally consists of digital circuits and analog circuits. However, it seems an aporia concerning the methods of the fault diagnosis and prognosis for analog circuits in current status. For real-time diagnosis and prognosis, various suitable intelligent algorithms are needed, and appropriate data is also needed to select and evaluate the intelligent algorithms. Since the real data are not easy to obtain, instead, simulation data can be considered. In addition, hard faults like open or short circuits are not usual but the gradual failures such as parameters shifting, usually happens. So, it is more valuable to simulate gradual failures and get their data for detecting the applicability of various diagnostic and prognosis algorithms. A voltage-controlled function generator, which wholly consists of analog circuits, is used for the case study in this chapter. First one or more resistors as the coming failure components are selected and defined according to the circuit diagram in the sensitive areas which are determined by thermal analysis, vibration analysis and life analysis. Then, a set of simulation data at different test points is obtained by the Multisim software through gradually varying resistance values of the defined failure resistors while simulating the tolerances of the other normal components by Monte Carlo method at the same time. At last, an experiment was carried out to prove the feasibility of the above simulation method by using the real rheostats instead of the defined failure resistors, and a set of test data at different test points is obtained.

Keywords Gradual failures · Analog circuits · Diagnosis · Prognosis · Simulation

S. Xie (✉) · S. Zhang · J. Liu
National Laboratory for Reliability and Environmental Engineering, School of Reliability and Systems Engineering, Beihang University, Beijing, China
e-mail: yeerxs@dse.buaa.edu.cn

S. Zhang
e-mail: zsn@buaa.edu.cn

J. Liu
e-mail: liujiaming@dse.buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_140

1 Introduction

Prognostic and Health Management (PHM) is a key technology for condition-based maintenance and autonomic support, and electronic systems are one of the main application areas in PHM [1]. However, it seems an aporia concerning the methods of the fault diagnosis and prognosis for analog circuits in current status. For real-time diagnosis and prognosis, various suitable intelligent algorithms are needed, and appropriate data are also needed to select and evaluate the intelligent algorithms. Since the real data are not easy to obtain, instead, simulation data can be considered. Analog circuit failures can be divided into hard faults and gradual failures according to the fault degree. Hard faults are those the element parameters have a sudden and distinct change, like open circuits, short circuits and components damages [2]. Gradual failures refer to the parameter values deviate from the nominal values even beyond the tolerance range. Because hard fault simulation is more easily, a lot of research is based on hard faults at present. But hard faults such as open circuit and short circuit are not common circuit fault modes, and gradual faults based on some parameters shifting occur more easily and are common failure modes. So paying attention to gradual failures and their data is more practical for studying various diagnostic and prognostic algorithms.

There are a number of different methods of fault injection technique in practical applications, and it has formed the system theory. In existing researches, there are many ways to implementing fault injection and pin level-hardware fault injection method is the main way followed by the software implemented fault injection [3]. Although there have been many achievements in the research of fault injection at present and many useful tools are put into use, however a unified and formal model of fault injection technique is lack of research. Most research has focused on two aspects: (1) the improvement and innovation of concrete realization method of fault injection [4]; (2) specially designed fault injection test in some specific projects [5]. Study limitations lead to specific fault injection technique in the study simply instead of system application. One of the developing directions of fault injection technique is to combine fault injection with all kinds of models and methods such as field measurement, combine fault injection with formal methods of new technology, combine fault injection used in the design phase with fault injection used in evaluation stage and combine fault injection for hardware assessment with fault injection for software evaluation [6]. This article is combining actual circuits and the simulation software and using the simulation methods to realize fault injection for analog circuits.

For imitating circuit system faults, a lot of circuit simulation software has the corresponding fault injection methods, in Multisim software, for example, a resistance's hard fault simulation can be realized directly in the resistance element attribute editor of the software, however, all the existing simulation software cannot directly realize gradual faults. So far, there has been not a fault injection method and theory for gradual faults recognized by the industry.

This chapter, for the realization of the gradual failures, designs a set of corresponding fault injection method, that is, according to the circuit principles, working

conditions and circuit faults that may occur, we set component parameter curves in the circuit faults in advance. Then we replace the time cumulative effect with the simulation number and successively simulate according to the parameters change. By this way, it can be more real reflect how the fault happened in actual working conditions. And we can obtain the performance degradation simulation data that can show how the actual circuit works, and provide sufficient and reliable test data for the follow-up study of circuit fault diagnosis algorithm.

This chapter will take an analog circuit—voltage-controlled function generator circuit [7] as an example, get a set of available failure data by the simulation method, and provide relevant data for life prediction algorithms study. By this way, test cost should be saved in some extent.

2 Steps of the Simulations

The electrical system fault injection method and simulation data acquisition process of this chapter are shown in Fig 1. The specific steps are as follows:

(1) Principle analysis for the circuits: first of all is to analysis the working principle of the circuits and identify the logical relationship and signal flow process for the circuits. This step is the basis of circuit analysis and simulation.
(2) Life cycle load analysis: the main purpose is to find the main factors affecting the service life of the circuits. We need to find out the influence factors, such as temperature and vibration, etc., and to quantify these influence factors according to actual conditions. This step usually needs to combine the real condition of electronic products with the function of itself and refer to the experience and the design indexes.

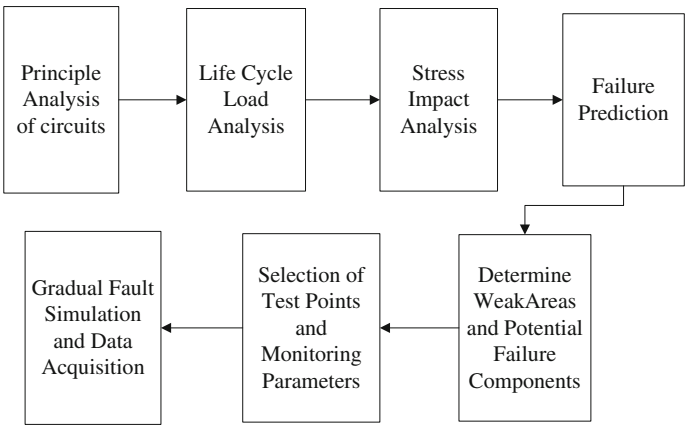


Fig. 1 Steps of Fault injections of gradual faults by the simulation methods

(3) Stress analysis: The common analysis is thermal stress analysis, vibration analysis, etc. They rely mainly on the FEA (Finite Element Analysis) simulation analysis software to complete the process which takes load and product information as inputs and gets stress distribution of the electronic systems as output.
(4) Fault prediction: this article adopts the PWA software that CACLE developed to take out fault prediction analysis on the circuit board. It is a kind of method based on physics of failure. It mainly helps to find out the weak areas of the whole electronic system and potential failure components.
(5) Determine the weak areas and potential failure components: According to the results of the stress analysis and fault prediction analysis, and based on circuit principles, and also considering the need of products to complete a specific function, we can find out the weak areas and potential failure components. Under normal circumstances, the weak areas are mainly located in the high temperature areas, large vibration response areas or short-life components.
(6) The selections of test points and monitoring parameters: After making sure the weak areas and potential failure components, we need to set up monitoring sites that can effectively reflect the changes of the circuit performances.
(7) Gradual fault simulations and data acquisitions: Using the simulation software, the gradual fault simulations will be conducted to obtain the simulation data.
(8) Application of simulation data: The simulation data obtained can be used for evaluating or developing life prediction algorithms and guiding electronic system experiments, etc.

3 A Case Study

3.1 Principle Analysis for the Circuits

A voltage-controlled function generator, which wholly consists of analog circuits, is used for the case study in this chapter. Figure 2 shows the circuit diagrams of its each modules, then the whole PCB diagram are drawn after selecting appropriate real components, as shown in Fig. 3 which can be seen the layout of the components of the circuits.

The circuit is controlled by the voltage to generate square wave, triangle wave and Sine wave. Changing the controlled voltage Uc, the frequency of the three kinds of output waveform can be changed.

3.2 Life Cycle Load Analysis

The voltage-controlled function generator generally works at an ambient temperature cycle of 25–70 °C. So the environmental temperature load throughout life

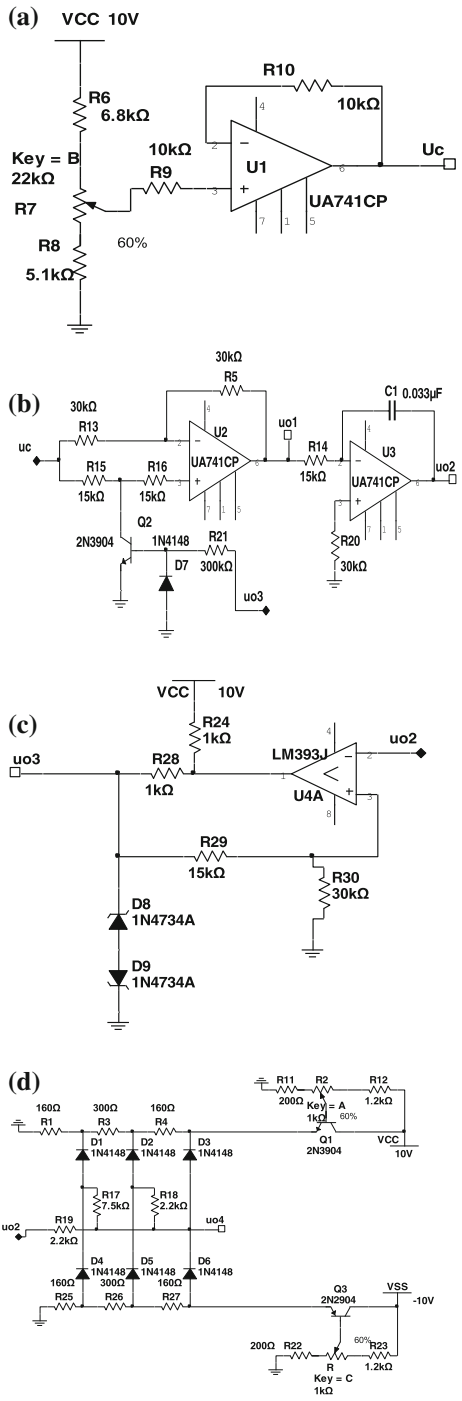


Fig. 2 Circuit diagrams of the voltage-controlled function generator. **a** The module of voltage controlling, **b** The module of square wave generating, **c** The module of triangle wave generating, **d** The module of Sine wave generating

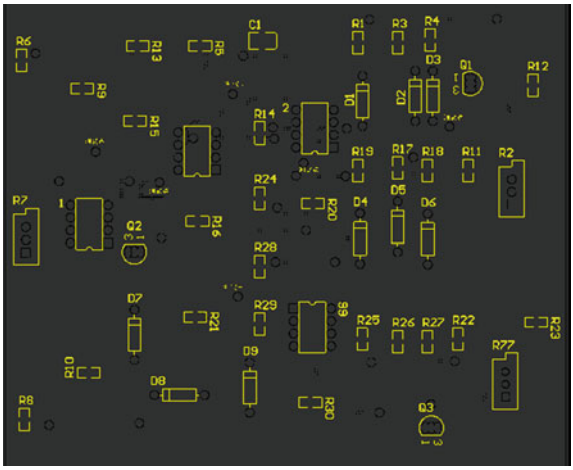


Fig. 3 A PCB diagram of the voltage-controlled function generator

cycle of the voltage-controlled function generator is cycling from 25 to 70 °C.The rates of change from 25 to 70 °C and from 70 to 25 are both 3 °C/min and the durations at lowest and highest are both 60 min. In addition, there is an assumption that the equipment continues to work everyday. The voltage-controlled function generator, as one of the measuring instruments, is often placed in a relatively flat and stable environment. So vibration, corrosion, and other loads can be ignored.

3.3 Stress Analysis

After summarizing and simplifying life cycle loads, the next step is to analyse its temperature stress. We choose the Flotherm (9.2) as the thermal stress analysis software. Flotherm software is a simulation analysis software for electronic system thermal dissipation, which developed by the British FLOMERICS Software Company. The software uses a sophisticated CFD (Computational Fluid Dynamic) model and numerical simulation technology, and successfully combines the experiences and databases of the FLOMERICS Company on thermal transfer of a large number of electronic devices, and also, Flotherm software has a large number of developed specifically model libraries for the electronics industry.

The CAD model of the voltage-controlled function generators input to the Flotherm 9.2. The input and output of the thermal stress analysis is shown in Table 1 [8]. The components whose temperature exceeds the allowable temperatures are regarded as thermal sensitive areas. The thermal analysis results are shown in Fig. 4. It is observed that, the component colour is nearer red, the temperature of the component is higher.

Table 1 Input and output of the thermal stress analysis

Input	Output
CAD model of the electronic system	Overall temperature distribution of the electronic system
Temperature loads in the life cycle; Cooling ways of the product	Temperature distribution of the PCB board and components
Designing power consumption of every level of the system;	Components whose temperature are over allowable temperature
Power consumptions by functional simulation	……
Material properties	

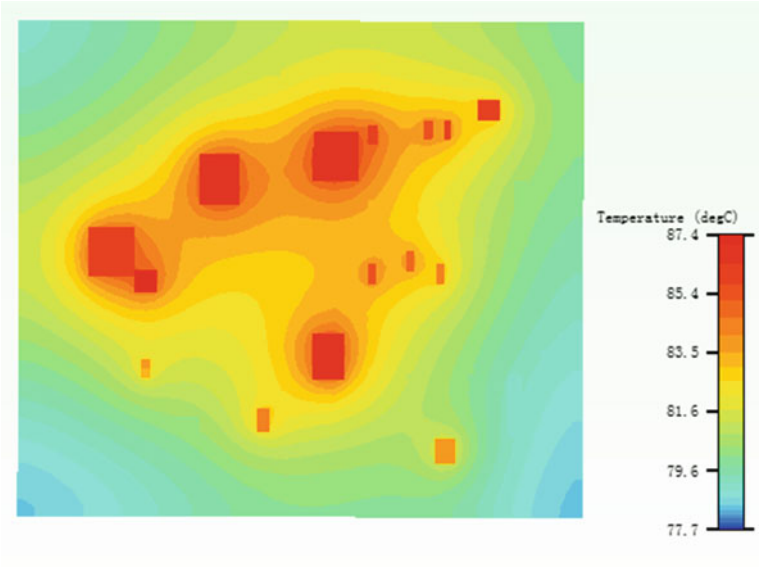


Fig. 4 The result of the thermal analysis

3.4 Fault Prediction

Based on the thermal analysis results, a fault prediction for the circuit board is carried out using the PWA software developed by the CACLE at the University of Maryland.

A detailed component distribution model, circuit board materials information, electrical parameters, shape parameters, working environments and thermal stress analysis results of the circuit board are input to the PWA software. Matched appropriate failure physical models for the components of the circuit board, the time to failure for each component is calculated, as shown in Fig 5. Figure 5 shows the relatively weak areas in the yellow cross, and all weak components are resistance.

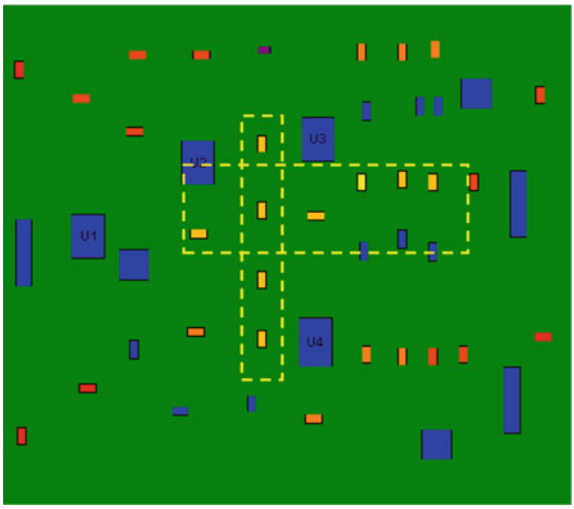


Fig. 5 The result of the fault prediction

3.5 Determine the Weak Areas and Potential Failure Components

It can be seen from Fig. 4 that the highest temperature area is the area around the four integrated chips (3 UA741 and 1 LM393). This is because that the maximum power consumption is from the chip under normal working condition.

It can be seen from Fig. 5 that short-life components are the SMD resistors around the chips. The first reason is that it has fully considered the thermal dissipation in chip encapsulation process although the chip temperature is higher in use. The second reason is that the life of SMD package is shorter than the life of the in-line package by matching the thermal solder joint fatigue model in the PWA software.

R19 plays an important role when triangle wave generates Sine wave. The resistor connects the input circuits and output circuits. The slope and the peak of the triangle wave flowing through the resistor R19 influent the shape of the Sine wave. R19 locates in the weakest area from the results of fault prediction (yellow cross area). Summarizing the above reason, R19 which is the potential failure component is selected as fault simulation in the case study.

3.6 The Selections of Test Points and Monitoring Parameters

Six test points are selected for simulation, as shown in Fig. 6. The reason to choose the six test points is as follow. The outputs of all the test points are voltage for convenience. In terms of circuit principles, TP3, TP5 and TP6 are the sites for

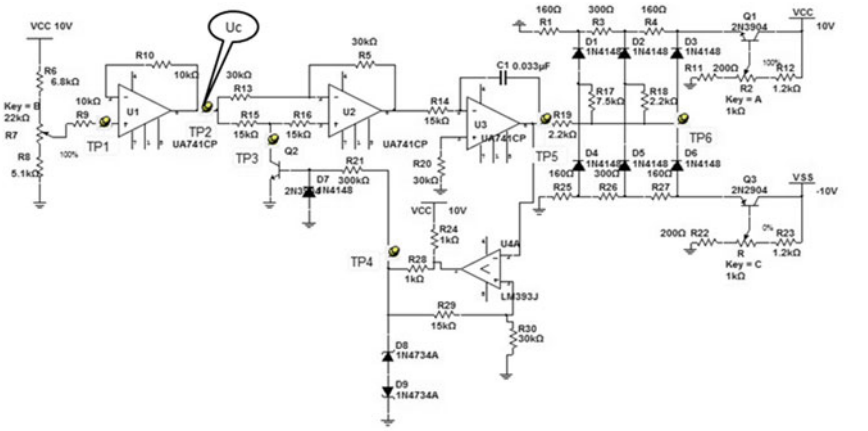


Fig. 6 Test points in the circuits

outputting square wave, triangle wave and Sine wave, and their outputs are equivalent to the final outputs and the evaluation indexes of the circuit performances. TP1, TP2 are used to testing the control voltage Uc, so they can isolate the fault resistance in front. The amplitudes of the three kinds of waveforms are controlled by the D8 and D9 stabilivolts and TP4 can monitor the two components, so TP4 can detect if waveform amplitude is stable.

Because the frequencies and amplitudes of the test point voltages are not all the same. In order to normalize data, the effective value of output voltage of each test point is selected as the final data. In equal period of time, if a DC voltage generates the equal heat through the same resistance generated by an AC voltage. The DC voltage can be called the effective value of the AC voltage.

3.7 Gradual Fault Simulations and Data Acquisitions

3.7.1 The Selection of Simulation Software

The Multisim 12.0 is selected, which is a Windows based simulation tool issued by the United States National Instruments (NI) co., LTD, and is applicable to the board-level design work for analog/digital circuit. It has strong ability for simulation analysis with the input ways including graph inputs of a circuit diagram and description language inputs of circuit hardware, and is suitable for low frequency analog circuit and digital logic circuits There are many ways to read out the value of the monitoring parameters, including using the oscilloscopes or multimeters and other common virtual instruments directly to get data, using dynamic probe points to read data and using the output charts to read and record data. In this study we select to connect a virtual multimeter to each test point to read data at every

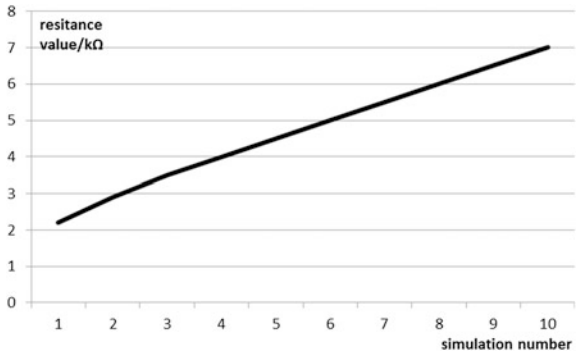


Fig. 7 The parameter shifting curve of R19

Table 2 Resistance changes of R19

Simulation number	1	2	3	4	5	6	7	8	9	10
Resistance value/KΩ	2.2	3	3.5	4	4.5	5	5.5	6	6.5	7

simulation time. So that we can conveniently and effectively read data and the real-time data display comes true. In addition, this method also can store data by using the storage capability of the software.

3.7.2 Failures Setting Beforehand

Figure 7 and Table 2 show the parameter drifting curve and corresponding resistance values of R19 whose nominal value is 2.2 KΩ. The resistance value increasing gradually imitates a fault mode of solder break caused by temperature cycling.

3.7.3 Monte-Carlo Simulation

Monte-Carlo analysis is a statistical analysis method, which exist in common circuit simulation software giving the statistical distributions of the tolerances of component parameters, we can use a set of pseudo random numbers to obtain the random sampling of the components. We can analysis these random sampling circuits and estimate the statistical distribution of circuit performance through the analysis results.

In this study, the tolerances which equal to 5 % nominal resistance values for all resistances except R19 are set by selecting "use tolerance" option under the menu of Multisim 12.

The first method to take Monte-Carlo is that the simulation of Monte-Carlo analysis can be found under the menu option. Then we can define Monte-Carlo

Table 3 The data of all test points on normal condition

Simulation number	TP1	TP2	TP3	TP4	TP5	TP6
1	5.007	5.008	2.429	6.028	5.039	3.82
2	4.977	4.978	2.413	6.026	4.954	3.745
3	5.019	5.02	2.438	6.029	5.026	3.771
4	4.995	4.996	2.422	6.024	4.958	3.783
5	4.991	4.992	2.418	6.027	5.031	3.833
6	5.017	5.017	2.431	6.027	5.127	3.867
7	5.025	5.026	2.441	6.028	5.068	3.895
8	4.977	4.978	2.415	6.028	5.026	3.839
9	4.984	4.984	2.422	6.029	4.977	3.803
10	5.02	5.021	2.437	6.028	5.044	3.779

parameters according to the need which are type and size of the tolerance, number of simulation, monitoring parameters and circuit working condition. The deficiency of this method is it can only output one monitor parameter at the same time. This method can't achieve to output multiple monitoring parameters.

The second way is that there is a "use tolerance" option under the menu. Choosing it, every simulation will give all the monitor parameters. But it can only manually simulate. This chapter had chosen this way to carry out Monte-Carlo analysis.

Firstly the output data of each test point under normal condition are measured so as to be compared with the data obtained through fault injection. Table 3 shows the data of all test points on normal condition, and the number of Monte-Carlo simulation is 10.

It can be seen from the data in the table that only TP6 can completely isolate the faults of R19. This is related to the position and function of R19. Of course, the premise of this is that other faults do not occur. From the previous stress analysis, we know that R19 is a relatively weak component in the whole circuit. So if the data monitored is similar to the data in Table 4, although we can't rule out if other fault resistances exist, it is the largest undoubtedly that R19 has breakdown.

4 Conclusion and Future Work

In this chapter, we use Multisim 12.0 which is commonly used simulation software to realize the gradual fault occurring on the selected circuit. By using simulation times instead time, a solder break simulations are conducted according to a curve set in advance, which the resistance of a resistor varies with the simulation time gradually increasing until opening, while the Monte Carlo analysis on other resistors in the circuit by setting random variation tolerances are conducted to obtain a set of more realistic simulation data, which can provide for algorithms of

Table 4 Some fault data when R19 is imitated gradual faults

	TP1	TP2	TP3	TP4	TP5	TP6
2.2 K	5.007	5.008	2.429	6.028	5.039	3.82
	4.977	4.978	2.413	6.026	4.954	3.745
	5.019	5.02	2.438	6.029	5.026	3.771
	4.995	4.996	2.422	6.024	4.958	3.783
	4.991	4.992	2.418	6.027	5.031	3.833
	5.017	5.017	2.431	6.027	5.127	3.867
	5.025	5.026	2.441	6.028	5.068	3.895
	4.977	4.978	2.415	6.028	5.026	3.839
	4.984	4.984	2.422	6.029	4.977	3.803
	5.02	5.021	2.437	6.028	5.044	3.779
3.5 K (+50 %)	4.958	4.959	2.406	6.025	5.027	3.563
	4.975	4.975	2.414	6.027	5.028	3.55
	4.946	4.947	2.402	6.027	5.102	3.55
	4.985	4.986	2.418	6.027	5.012	3.556
	5.017	5.017	2.435	6.028	5.088	3.54
	4.96	4.96	2.409	6.025	5.033	3.548
	5.005	5.006	2.431	6.027	5.016	3.542
	4.994	4.995	2.423	6.028	5.04	3.52
	4.984	4.984	2.413	6.026	5.01	3.53
	5.003	5.004	2.428	6.029	5.108	3.58
4.5 K (+100 %)	4.969	4.97	2.415	6.027	5.052	3.34
	5.023	5.023	2.443	6.026	5.066	3.333
	4.961	4.961	2.41	6.028	5.05	3.351
	5.033	5.033	2.442	6.028	5.052	3.316
	5.093	5.094	2.474	6.032	5.135	3.329
	5.006	5.006	2.431	6.028	5.019	3.316
	4.983	4.983	2.42	6.027	5.061	3.345
	5.063	5.063	2.458	6.03	5.064	3.324
	5.032	5.033	2.444	6.029	5.051	3.339
	5.055	5.056	2.451	6.025	5.022	3.306
5.5 K (+150 %)	5.03	5.031	2.448	6.025	5.78	3.12
	4.986	4.987	2.421	6.026	5.073	3.183
	4.971	4.971	2.414	6.025	5.041	3.174
	4.994	4.995	2.422	6.029	5.082	3.184
	5.065	5.065	2.459	6.029	5.079	3.149
	4.994	4.995	2.426	6.026	5.035	3.164
	5.04	5.041	2.488	6.027	4.997	3.125
	4.937	4.938	2.399	6.026	5.048	3.164
	4.998	4.998	2.424	6.026	5.042	3.165
	5.073	5.073	2.463	6.026	5.028	3.172

diagnosis or prognosis. This simulation program is a reasonable, simple and effective method for fault injection and fault simulation.

Following aspects should be further studied:

(1) The Monte Carlo simulation method in Multisim 12.0 can be only used on a voltage node if using the built-in function,and another way is manual simulation after making sure all test points. The former method can't get simulation data of multiple test points at the same time, however, the latter approach needs the manual simulation and it will cost much time if we need a lot of simulation data. So it urgently needs to find a reasonable solution.
(2) In this chapter, fault resistance of the circuit we chose has a representative meaning, but the selection of test points is not ideal. It can be seen that test point does not adequately reflect the changes of R19 to the overall impacts on the circuit by observing the simulation data. So the selection of test points needs to further research.
(3) In this study, we chose the resistor as the fault component, however, in other circuits, the components which are easily breakdown may be capacitors, chips, transistors, etc. As long as we analyze the principle of fault components and make clear which parameter changes, the method in this chapter can be used. The difference is the curve of parameters.

References

1. Pecht M, Kang R (2010) Fault diagnosis, prediction and system health management. PHM Centre, City University of Hong Kong
2. Zou R(1989) Principle and method of analog circuit fault diagnosis. Huazhong University of Science Press, Huazhong
3. Sun JC, Wang JY, Yang XZ (2001) The current research status of fault injection method and tool. J Aerosp 22(1) (in Chinese)
4. Amendoia AM, Impagliazzo L, Marmo P et al (1997) Experimental evaluation of computer-based railway control systems. In: Proceedings on 27th IEEE international symposium on fault tolerant computing (Ftcs227), Seattle, pp 380–384. doi:10.1109/FTCS.1997.614112
5. Stott DT, Ries G, Hsueh MC et al (1998) Dependability analysis of a high-speed network using software implemented fault injection and simulated fault injection. IEEE Traps Comput 47 (1):108–119. doi:10.1109/12.656094
6. Ye YF (2008) Software fault injection method based on the model. Central China Normal University, Wuhan
7. Zhang YP (2008) Electronic technology experiment. Beijing Institute of Technology Press, Beijing, pp 78–85
8. Wang XF, Li ZQ, Zhnag SN, Liu JM, Shao C (2013) Test point selection based on functional simulation and FMMEA for an electronic system on PHM

Career Employability Development Through a Specialized Asset Management's Degree: An Exploratory Analysis for a Chilean Program

Edward Johns, Raúl Stegmaier, Jorge Cea, Fredy Kristjanpoller and Pablo Viveros

Abstract Asset Management as a growing discipline is being incorporated into the Chilean Industry, especially into the mining Industry; however, there is a need to provide the necessary skills and knowledge regarding these new positions. In that regard, since 2005 we have successfully incorporated a master's program scheme entitled Master of Asset Management (MGA, currently in its 9th offering). Up to date, more than 300 postgraduate students from different industries have attended this MGA, becoming the leading program in Latin America. In the training field we have developed a Diploma on Asset Management (DGA—140 h). With the present study, we have analysed the students' profiles to identify the industry distribution and, at the same time, to measure the extent to which the skilled labour (graduates) have moved between different job positions and/or companies, be it as a horizontal mobility (does not result in a change in the worker's grading or status) or a vertical mobility (if it does). Additionally, we regularly test the influence of the market-value and scholastic and social dimensions of the human capital on the alumni. The scope is mainly the Chilean industry.

E. Johns (✉)
Asset Management and Maintenance, Department of Industries, Universidad Santa María, Av. España 680, Valparaiso, Chile
e-mail: edward.johns@usm.cl

R. Stegmaier
Department of Industries, Universidad Santa María, Valparaiso, Chile
e-mail: raul.stegmaier@usm.cl

J. Cea · F. Kristjanpoller · P. Viveros
Universidad Santa María, Valparaiso, Chile
e-mail: jorge.cea@usm.cl

F. Kristjanpoller
e-mail: fredy.kristjanpoller@usm.cl

P. Viveros
e-mail: pablo.viveros@usm.cl

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_141

1 Introduction

Entering a postgraduate program helps to increase knowledge and skills in specific areas of a professional career and also allows a higher personal development, which can be a key element in career success [1] and higher employability [2]. The decision for enrolling in a Master's Degree often implies investing a large sum of money, time and effort; thus, getting to know the potential return such a personal project may yield in the future would be incredibly useful for prospective students, when comparing these with other possibilities.

Baruch et al. [4], have put forth a methodology that allows for analysis of competences, skills, human and professional capital development, recognized upon completion of a postgraduate degree. By surveying alumni from a US general MBA and specialist Master's program, they were able to measure the *added value* given to a professional's life through the postgraduate studies.

Based upon this methodology, a similar study is performed on the Professional Master of Science in Maintenance Engineering and Asset Management (MGA, in its abbreviated Spanish form: Magíster en Gestión de Activos y Mantenimiento), a specialized Master's degree primarily focused, as its name says, on Asset Management and Maintenance Engineering, which has been offered by the Departamento de Industrias of the Universidad Técnica Federico Santa María (UTFSM), Chile. To date, 9 generations of professionals have taken it as part of their professional training since it opened in 2005, as a pioneer postgraduate degree and as a leader in the field in Latin America. This 2-year program is comprised of 400 academic hours in a weekend schedule format.

With approximately 330 alumni and 210 graduate students to date, this program's first generation has already been active in the market for about 6 years, allowing the main objective of this investigation: what effect an Asset Management degree, such as the MGA, has had on the training, the employability and professional development of its graduates.

2 Theoretical Framework

Postgraduate programs offer their students in-depth and specific knowledge and a more polished set of skills for certain professional areas, new tools and a renewed chance to approach a university's learning setting. This experience nurtures the career and life of those who pursue such a degree, by improving their competitiveness, employability [3] remunerations, and also an easier entrance into a valuable professional network.

As discussed and researched on [4], the human capital, upon which a graduate's program may have effect, can be analysed in five dimensions: scholastic capital (referring to the knowledge acquired on a degree); social capital (indicating the creation of valuable networks); cultural capital (the value that society assigns on

prestige symbols); inner-value capital (the managerial skills gained through a higher sense of self-esteem, self-awareness and confidence); and, lastly, market-value capital, which depends on the previous four, and is evidenced by the higher remuneration graduate students get.

A good postgraduate degree must give valuable and relevant knowledge and information to its students, distinguishing them from the rest of the professionals available in the market; and thus improving their employability assets by increasing the *scholastic capital* owned.

Social capital is also increased through participation in the networks generated in postgraduate's degrees. The value-added element in these networks ultimately promotes the success of the organization where the graduate works [5]; thus becoming an asset for both the individual and the company.

Attaining a higher degree of education helps to climb the social status ladder, gain prestige and a higher recognition by society. A direct correlation between MBA and specialized programs and this positive impact on the *cultural capital* has been established in previous studies (see [6, 7]).

Inner-value capital considers the internal competences upon which a person builds self-image and their improvement represents a higher level of performance [8], which is sought after by companies.

In the case of the MGA Master's degree, of these five aspects of human capital, only three are taken into account for an exploratory analysis: scholastic capital, social capital and market-value capital. It has been previously analysed how MBA alumni have seen an increase in their wages [9].

A broader vision on the employability assets that a specialized degree may bestow on its students for competing in the labour market is thus given.

Based on this theoretical framework three dimensions were taken into account which has led to three research hypotheses, as follows.

Hypothesis 1 The Professional Master of Science in Maintenance Engineering and Asset Management degree has had a positive influence in market-value capital of its graduate students.

Hypothesis 2 The Professional Master of Science in Maintenance Engineering and Asset Management degree has had a positive influence in the scholastic dimension of the human capital of its graduate students.

Hypothesis 3 The Professional Master of Science in Maintenance Engineering and Asset Management degree has had a positive influence in the social dimension of the human capital of its graduate students.

3 Methodology

3.1 Research Delimitation

Primary objectives of this investigation are the definition and limitation of the effects the Master of Asset Management has had on the career of its graduates, i.e., to specify which area is of particular interest when studying the professional development of the alumni, by observing the changes they have gone through from the time of enrolment, graduation and current position. In order to do so, an appropriate methodology able to measure these implications must be found.

The first considered evidence is the possibility of work position movement in a company, to another company or to another industrial field. This possibility given at some point after the completion of the degree may have been gained thanks to it. Therefore, it is important to find out if the MGA had any such influences. Along with this, the chance of rising in the hierarchical ladder and gaining higher levels of responsibility are also analysed, to see if the work position change also involved a vertical shift in the organizational chart.

Another important bit of evidence, helpful when measuring the return from investment on a Master's degree program, is the wage difference between what was earned before the program, right after its conclusion and in the long term. This shows how the market reacts to an improvement in the professional assets of a worker that has gone through a postgraduate program.

Lastly, the scholastic and social capitals are surveyed in an exploratory manner, in order to assess the development perceived by the alumni.

3.2 Sampling and Survey

Once the areas to study are clarified, a brief and anonymous on-line survey is created to evaluate their career development, consisting of 25 questions in 8 groups of similar content. During June of 2013, the on-line examination was sent to 258 ex-students, whose e-mails were available, updated and correct. The entire population of ex-students consists of 337 people, resulting in a gap between the potential sample and the actual population sample of 23.4 %.

Following the initial mailing, and three consecutive reminders, 84 completed surveys were retrieved, representing a 32.6 % response rate, consistent with previous studies, as it appears on [4]. All the answers were given by male graduates (only 10 women have graduated so far), with about 80 % of them ranging in age from 30 to 50 years old. Participants have worked in an array of industries, but most of them (62.2 %) in the mining sector. This bias is explained by the nature of Chilean industry whose primary source of employment is related to copper, being Chile the world's largest producer of this mineral.

In the survey, one question was focused on determining what generation the survey represents, given that the program has been offered since 2005. The time

between graduation and the moment of evaluation is relevant. As a demographics measurement, there were 2 questions focused on defining the age and gender of the graduate.

A set of direct questions is presented to determine the industrial sectors in which the graduate has worked before the MGA, right after it was completed and presently; thus, knowing the employment *path* taken by the graduate. A list of the 11 most common industries among MGA participants was obtained from a previous self-evaluation of the program [10], being, for instance, copper mining, engineering services, transportation and telecommunication. A blank box was offered, so that the graduate could write a sector not previously considered.

The perception regarding the hierarchical position shift occurred since the attainment of the degree until now is asked through a direct question. It is set by placing statements over a personal opinion, such as "I've risen by 2 hierarchical levels", "I'm on the same position" or "I've fallen 1 hierarchical level" between the enrolment and the conclusion of my degree, and between graduation and my current status. This resulted in a 7 level possibility response, from "I've fallen more than 2 levels" to "I've risen more than 2 levels". This is based upon the realization of the existence of complex organizational structures in certain industries, in which it is sometimes hard to tell where exactly a certain hierarchical position is located on the organizational chart. It also allows a comparison among different industries and companies, like retail, mining and food process.

In addition to this, an open question asking for the name of the position was given, so it could be observed what kind of work the professional has had before and after graduation from the Master's degree, e.g., execution, supervision or control.

As a measurement of the level of responsibility a worker has had, it is asked how many people the graduate had under direct responsibility, before entering the MGA, after its completion and currently. The response possibilities are as follows: "0 people", "1–5", "6–10", and so on, up to "More than 25".

Development of the scholastic and social capitals are analysed through 6 questions. In these questions, 4 level Likert response possibilities are given to positive statements, i.e., "The program helped me improve my skills on the asset management field", "In the Master's degree I acquired new knowledge focused on reliability processes in my work", "Thanks to the MGA I could create a valuable network from the people I met" or "I've used the network, exchanging useful information". The 4 response possibilities are the following: "I disagree completely", "I disagree", "I agree" and "I agree completely".

4 Results

Participants of the survey showed a mean of 3.74 years (SD 1.64) of work life after graduation from the Master on Asset Management. Table 1 shows details regarding the number of responses per generation.

Table 1 Number of responses per year of graduation

Graduation year	2006	2007	2008	2009	2010	2011	2012	Total
Nr of responses	5	11	10	18	14	24	2	84

The largest number of responses was given by ex-students from the generation of 2009, 2010 and 2011, with a 66.67 % response rate.

Hierarchical position change possibilities were assigned levels to quantify the received responses, i.e., "*I've fallen more than 2 levels*" was given a −3, "*I've fallen 2 levels*" was given a −2, and so on, up to "*I've risen more than 2 levels*", that was given a +3. Furthermore, the time between enrolment and graduation was named "*1st stage*", and the time between graduation and the current situation was labeled as "*2nd stage*". Relative response frequencies to this question for 2008–2011 generations are given in Table 2. As it can be seen in the highlighted statistical mode for every generation and moment, most data occurs above Level 0 (same hierarchical position), so it is safe to assume that on average, a graduate from the MGA would see their position safe, in terms of not descending in their organizational chart.

Table 3 shows one-variable statistics for the 2008–2011 generations. The overall change represents the 'leap' since enrolment in the MGA program. Generations that graduated earlier (2008 and 2009) have a higher change in their position than the most current ones.

Responsibility degrees, in number of people under a worker's responsibility, were also assigned to levels, as in the previous analysis, to quantify the types of responses received. In this case, the responses were given levels from 0 to 6, i.e., "0 people" was given a 0, "1–5 people" was given a 1, and so on, up to "More than 25" with a 6. Table 4 shows the relative response frequency given to this question by the graduates.

Table 2 Relative response frequency for hierarchical level change

Grad. year	2008		2009		2010		2011	
Resp. level	1st stg. (%)	2nd stg. (%)	1st stg. (%)	2nd stg. (%)	1st stg. (%)	2nd stg. (%)	1st stg. (%)	2nd stg (%).
Level 3	*30.00*	20.00	11.11	22.22	7.14	7.14	4.17	0.00
Level 2	10.00	10.00	22.22	11.11	7.14	14.29	0.00	8.33
Level 1	20.00	*40.00*	22.22	*33.33*	*42.86*	14.29	*58.33*	29.17
Level 0	*30.00*	20.00	*38.89*	27.78	*42.86*	*64.29*	33.33	*62.50*
Level-1	10.00	10.00	5.56	5.56	0.00	0.00	0.00	0.00
Level-2	0.00	0.00	0.00	0.00	0.00	0.00	4.17	0.00
Level-3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total	100	100	100	100	100	100	100	100

Table 3 One-variable statistics for hierarchical position change

Hierarchical position change	Mean			
Graduation year	2008	2009	2010	2011
1st stage	1.09	0.94	0.73	0.62
2nd stage	1.00	1.16	0.60	0.45
Overall change	2.09	2.10	1.33	1.07

Table 4 Relative response frequency for responsibility levels, generation and time of interest

Gen.	Time[a]	0 (%)	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)	Total (%)
2008	A	20.00	20.00	20.00	0.00	0.00	0.00	*40.00*	100
	B	0.00	10.00	0.00	10.00	20.00	0.00	*60.00*	100
	C	10.00	0.00	20.00	0.00	0.00	0.00	*70.00*	100
2009	A	5.56	*27.78*	0.00	22.22	22.22	11.11	11.11	100
	B	5.56	16.67	5.56	*22.22*	5.56	*22.22*	*22.22*	100
	C	22.22	5.56	11.11	0.00	5.56	22.22	*33.33*	100
2010	A	21.43	*28.57*	14.29	14.29	7.14	0.00	14.29	100
	B	14.29	*28.57*	14.29	7.14	0.00	7.14	*28.57*	100
	C	21.43	21.43	*28.57*	0.00	0.00	0.00	*28.57*	100
2011	A	16.67	*29.17*	12.50	4.17	16.67	4.17	16.67	100
	B	12.50	*20.83*	*20.83*	4.17	4.17	*20.83*	16.67	100
	C	16.67	*25.00*	20.83	8.33	4.17	8.33	16.67	100

[a] *A* Before enrolling, *B* Right after graduation, *C* Current status

A relative frequency, observed in Table 4, gives each response from generations 2008–2011 a responsibility level, as previously described. As roughly seen, there seems to be an increase in the responsibility level, shown by the highest frequency on each response highlighted.

Tables 5 and 6 show one-variable statistics for responsibility levels and the changes, which occurred, respectively, for the generations graduated in 2008–2011.

The generation of 2008 has undergone the largest responsibility level change as they have the longest time in the 'labour market' since graduation. It is noticeable, that the generations of 2010 and 2011 have suffered a decrease in their responsibility levels, mainly in the '2nd stage' of their career, i.e., since the graduation.

Table 5 One-variable statistics for responsibility levels

Responsibility degree	Mean			
Graduation year	2008	2009	2010	2011
Before enrolling	3.00	3.06	2.07	2.54
Right after graduation	4.80	3.61	2.87	2.96
Current status	4.60	3.61	2.40	2.50

Table 6 One-variable statistics for change in responsibility level

Responsibility degree change	Mean			
Graduation year	2008	2009	2010	2011
1st stage	1.80	0.55	0.76	0.38
2nd stage	−0.20	0.00	−0.43	−0.42
Overall change	1.60	0.55	0.33	−0.04

Wages were analysed, by assigning a level to different wage intervals, as in the previous analyses. There are 10 different levels, from level 1 representing a monthly earning of less than $1,000,000 Chilean Pesos, around 2,000 USD (1 USD roughly equals 500 CLP [11]). From level 2 onwards, every interval represents an increase of $500,000 CLP (≈1,000 USD). Table 7 shows the relative response frequency for this analysis.

Table 8 shows one-variable statistics for the analyzed generations' responses, and there is, on average, a positive difference in salary, as seen on Table 9, during the whole process of attaining the Master's degree.

Figure 1 visually summarizes the information already given from the performance indicators analyzed. It can be seen how, on average, the graduates from the previous generations have attained higher degrees of 'development' in the areas of hierarchical position, income and responsibility level. Also, there is an overall positive change in each of the generations that has been studied with more detail, given that most of the indicators are positive. The negative change level in responsibility degree for the generation of 2011 is interesting to further analyze, but the limitations of the survey would not allow it.

Measurements on the results from the questions related to the development of scholastic and social capital returned the results shown on Table 10. Each possible response for this question was given a level, according to the agreement between the statement given and the perception the participant has had, regarding their experience with the MGA. "I disagree completely" was given a level of 1, and "I agree completely" was assigned to a level of 4. Thus, a mean close to 3 on both of these capitals represents a perception by the graduates that there has been progress in their personal assets through their participation in the MGA. Further details on the dimensions of human capital analysed are given on Tables 11 and 12.

It can be seen that scholastic capital has been developed at a greater degree than social capital. This happens, given the nature of the course. Students go twice a week to lectures and exercises, so the time is spent entirely on academic issues, rather than socializing.

Of the 84 completed surveys obtained, 21 people have changed at least once from an industrial sector to another. Upon observing the data acquired from the responses related to industrial sector movement, there is no evidence that this is in any way related to the remuneration difference or to the responsibility degree variation, previously analysed.

Table 7 Relative response frequency for wage levels, generation and time of interest

Gen.	Time[a]	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)	7 (%)	8 (%)	9 (%)	10 (%)	Total (%)
2008	A	20.00	*60.00*	0.00	20.00	0.00	0.00	0.00	0.00	0.00	0.00	100
	B	10.00	*30.00*	30.00	0.00	20.00	10.00	0.00	0.00	0.00	0.00	100
	C	10.00	*30.00*	10.00	0.00	10.00	10.00	*30.00*	0.00	0.00	0.00	100
2009	A	11.11	27.78	5.56	*38.89*	0.00	5.56	0.00	5.56	0.00	5.56	100
	B	0.00	*33.33*	0.00	27.78	11.11	11.11	0.00	5.56	5.56	5.56	100
	C	0.00	11.11	*16.67*	*16.67*	*16.67*	0.00	*16.67*	5.56	0.00	*16.67*	100
2010	A	0.00	50.00	0.00	*28.57*	7.14	7.14	7.14	0.00	0.00	0.00	100
	B	0.00	14.29	28.57	*35.71*	7.14	7.14	7.14	0.00	0.00	0.00	100
	C	7.14	0.00	*21.43*	*21.43*	*21.43*	7.14	*21.43*	0.00	0.00	0.00	100
2011	A	16.67	16.67	25.00	4.17	*29.17*	4.17	0.00	0.00	4.17	0.00	100
	B	0.00	13.04	*30.43*	17.39	13.04	17.39	4.35	0.00	4.35	0.00	100
	C	0.00	12.50	16.67	16.67	*20.83*	12.50	8.33	8.33	4.17	0.00	100

[a] *A* Before enrolling, *B* Right after graduation, *C* Current status

Table 8 One-variable statistics for income level (levels)

Income level	Mean			
Graduation year	2008	2009	2010	2011
Before enrolling	2.20	3.72	3.33	3.50
Right after graduation	3.20	4.50	3.73	4.26
Current status	4.20	5.50	4.46	4.83

Table 9 One-variable statistics for income level change (levels)

Income level change	Mean			
Graduation year	2008	2009	2010	2011
1st stage	1.00	0.78	0.40	0.83
2nd stage	1.00	1.00	0.73	0.52
Overall change	2.00	1.78	1.07	1.23

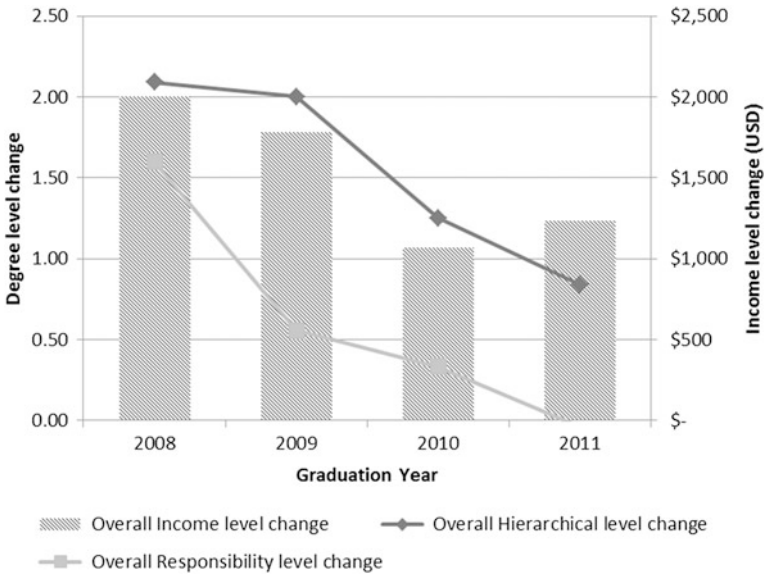


Fig. 1 Overall average change on performance indicators

However, when reading the job descriptions input, there seems to be a trend: from one position to a higher one on the organizational chart. As previously seen in the hierarchical position change analysis, this reinforces the positive movement perception in a qualitative manner. It is common to read an increase from "Assistant Manager" to "Manager in chief" or from "Unity chief" to "Section Head".

Table 10 One-variable statistics on perceived human capital development

Human capital	Mean
Scholastic capital	3.57
Social capital	2.98

Table 11 One variable statistics on perceived scholastic capital development

Scholastic capital	Mean
Skills	3.62
Methodology	3.58
Knowledge	3.51

Table 12 One-var. statistics on perceived social capital development

Social capital	Mean
Creating networks	2.97
Making professional friends	3.17
Use of networks	2.80

5 Conclusions and Implications

On Tables 8 and 9 it can be seen that the income level increases as we move into the *older* graduation years. The same pattern can be seen on Fig. 1. There is enough evidence to confirm **Hypothesis 1**. In order to achieve a more precise impact of the master's degree at the income level, future research should define control groups. Another benchmark that would be interesting to compare is this result against a more generic master's program such as an MBA with the same industry mix. This could lead to future research in the area.

Regarding the scholastic capital mentioned in the **Hypothesis 2**, it can be seen in Table 11 that the overall sample of 84 alumni perceived, as a mean, an important contribution to their specific knowledge in the asset management discipline which was acquired through the master's degree MGA. According to Table 11—scholastic capital—the three dimensions scored greater than 3.5 on a 1–4 scale, thus we can confirm **Hypothesis 2**. These figures allow confirmation that this program is providing specific knowledge transfer to students pursuing the MGA master's degree.

Regarding the social capital mentioned in the **Hypothesis 3**, it can be seen on Table 12, that the overall sample of 84 alumni perceived an important contribution to creating valuable networks based on the master's degree. Even though perceived value approaches a mean of 3, on a 1–4 scale, there is still a gap to be analysed because there is room for improvement on how to strengthen the networking process. Based on these results we can confirm **Hypothesis 3**.

6 Notes

The Data Base and Project Evaluation [10] with the information were obtained from the MGA's administrative office.

A.1 7 Appendix

The translation of the on-line survey questions and response possibilities made during June of 2013 is shown hereafter.

1. When did you enroll in the Master of Asset Management?
2. When did you finish the courses of the Master of Asset Management?
3. In which industrial field were you working before enrolling in the MGA?
4. In which industrial field were you working upon completing the MGA?
5. In which industrial field are you currently working?
   (Possible responses for questions 3, 4 and 5: Engineering services—Mining—Electricity—Transport and/or Telecommunications—Forestry—Energy Generation—Chemical Processes—Retail—Sanitation—Software Development—Astronomy—Plus an open space for a response not considered).
6. Regarding the level in the organizational chart of your working place. Since enrollment until completion of the MGA, what statement better describes your situation?
7. Regarding the level in the organizational chart of your working place. Since completion of the MGA until your current place, what statement better describes your situation?
   (Statements given for questions 6 and 7: I've risen more than 2 levels—I've risen exactly 2 levels—I've risen 1 level—I'm in the same position—I've fallen 1 level—I've fallen exactly 2 levels—I've fallen more than 2 levels).
8. Please, give the name that better describes your work at the moment of enrollment into the MGA.
9. Please, give the name that better describes your work at the moment of completion of the MGA.
10. Please, give the name that better describes your current work.
11. How many people did you have under direct responsibility upon enrollment into the MGA?
12. How many people did you have under direct responsibility upon completion of the MGA?
13. How many people do you currently have under direct responsibility?
    (Possible responses for questions 11, 12 and 13: 0–1 to 5–6 to 10–11 to 15 – 16 to 20–21 to 25—More than 25).
14. Regarding your average monthly income, in what range was it when you enrolled in the MGA?

15. Regarding your average monthly income, in what range was it when you completed the MGA?
16. Regarding your average monthly income, in what range is it currently?
    (Ranges given for questions 14, 15 and 16: Less than 1,000,000 CLP—Between 1,000,001 CLP and 1,500,000 CLP—Between 1,500,001 CLP and 2,000,000 CLP—Between 2,000,001 CLP and 2,500,000 CLP—Between 2,500,001 CLP and 3,000,000 CLP—Between 3,000,001 CLP and 3,500,000 CLP—Between 3,500,001 CLP and 4,000,000 CLP—Between 4,000,001 CLP and 4,500,000 CLP—Between 4,500,001 CLP and 5,000,000—More than 5,000,000 CLP).
17. Regarding your personal experience with the master's program, the MGA helped me improve my asset management skills.
18. Regarding your personal experience with the master's program, the MGA gave me new methodologies in the asset management field.
19. Regarding your personal experience with the master's program, in the MGA I acquired more knowledge focused on plant reliability.
20. Regarding your personal experience with the master's program, in the MGA I acquired more knowledge focused on plant reliability.
21. Thanks to the MGA I could make a valuable contact network.
22. In the MGA I could meet good classmates and friends.
23. I have been making good use of the network created by exchanging useful information.
    (Likert scale possible responses for question 17 through 23: I disagree absolutely—I disagree—I agree—I agree absolutely).
24. What is your age?
25. Gender.

References

1. Hilgert A (1998) Professional development of women and the executive MBA. J Manage Dev 17(9):629–643
2. Baruch Y (2009) To MBA or not to MBA. Career Dev Int 14(4):388–406
3. Baruch Y, Leeming A (2001) The added value of MBA studies–graduates' perceptions. Pers Rev 30(5):589–602
4. Baruch Y, Bell MP, Gray D (2005) Generalist and specialist graduate business degrees: tangible and intangible value. J Vocat Behav 67(1):51–68
5. Nohria N, Eccles RG (1992) Face-to-face: making network organizations work. In: Nohria N, Eccles RG (eds) Networks and organizations. Harvard Business School Press, Boston, pp 288–308
6. Tafjel H (1981) Human groups and social categories: studies in social psychology. CUP Archive, Cambridge
7. Baruch Y, Peiperl M (2000) The impact of an MBA on graduate careers. Hum Resour Manage J 10(2):69–90
8. Gist ME, Mitchell TR (1992) Self-efficacy: a theoretical analysis of its determinants and malleability. Acad Manage Rev 17:183–212

9. Hegarty S (1996) Do MBAs lead to a better job and a bigger salary? Works Manage 49 (2):61–65
10. Formulario de Solicitud de Acreditación Programas de Postgrado MGA. Master of Asset Management Data base and Project Evaluation
11. Base de datos estadísticos: Tipo de cambio. Currencies daily data. Banco Central de Chile. http://si3.bcentral.cl/Siete/secure/cuadros/arboles.aspx (2013). Accessed 20 June 2013

PHM Collaborative Design in Aircrafts Based on Work Breakdown Structure

Ying Ma, Wenjin Zhang and Jie Meng

Abstract Prognostics and health management (PHM) technique has shifted the focus from traditional status monitoring to health assessment, and has become a critical technique in the development of new weapons and in autonomic logistics. However, the design process of the PHM system is not integrated properly in the design of the aircraft, which makes it impossible to achieve an optimal overall design. With this consideration, we introduce the concept of collaborative design into the PHM design. By applying the technique of work breakdown structure, the PHM system and the corresponding aircraft system are decomposed in three aspects: the object aspect, the development aspect and the project aspect. Then the design process of the PHM system and that of the aircraft are integrated based on the breakdown structure in a collaborative way. Then the PHM design in the aircraft can be optimized and the life cycle cost can be reduced.

1 Introduction

Recently, with the development of the computer science, artificial intelligence, microelectronics and MEMS technique, the system becomes more and more complex and the system tends to be designed in a more integrated way. Consequently, the diagnostic technique for weapons is changed from the traditional corrective maintenance and periodic maintenance to the on-condition maintenance that can cover all the main systems and critical components. A new concept, i.e., the prognostics and health management (PHM) technique, has been proposed by the US Army to deal with the advanced test, maintenance and management for the new generation of weapon equipment.

The key in the realization of the PHM technique lies in three points: (1) to use a minimal number of specialized sensors to collect the data and information about the

Y. Ma (✉) · W. Zhang · J. Meng
School of Reliability and Systems Engineering, Beihang University, Beijing, China
e-mail: maying@dse.buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_142

system, (2) to access, monitor and manage the system health status with intelligent algorithms and reference models based on these data and information, and (3) to propose proper advices on the condition-based maintenance policy based on the system health status [2, 3]. Compared with the traditional fault detecting techniques, PHM has distinct advantages. The failure can be predicted and the operators can be informed in advance, which improve the safety and reliability of the system. The maintenance is condition-based and improves the system maintainability and optimizes the design of the logistic system. At the same time, the condition-based maintenance can also reduce the lifetime cycle cost and achieve a better affordability. Therefore, PHM has become one of the key enabling techniques in the automatic logistics [8, 4].

As a technique that is continually developing, PHM is far from perfect in the system integration [11]. More work has to be done in the collaborative design of the PHM system and its "host" system. For the moment, the PHM design and the aircraft design are implemented separately, where the PHM design is often carried out after the aircraft has been designed. The overall optimal design will never achieved in such way, for example, if the aircraft has been designed without considering the PHM system, the PHM design will always be faced with many restrictions, such as the sensors allocation. With this consideration, the paper analyses the PHM system in a three-dimension view: the object dimension, the working dimension and the phase dimension. Then the object (the PHM system) and the working (including the management, design and test) are decomposed along the phase dimension based on the work breakdown structure (WBS) technique. After then, the collaborative design is introduced in the PHM system design and the aircraft design, which ultimately optimizes the aircraft support system and reduces the development cycle and the overall cost.

2 PHM Analysis on Three Dimensions

As the information resource for the automatic logistics system of the aircraft, the PHM system provides the ability of the instant in situ fault detection and isolation, fault prediction and health management. To realize these tasks, the design of the PHM system involves a lot of issues. Based on the structure and the operating features, the PHM design can be viewed in three different dimensions: the object dimension, the working dimension and the phase dimension, as shown in Fig. 1.

In the object dimension, the PHM system includes the on-board PHM system and the off-board PHM equipment. With the hierarchical reasoning and the integrated fault diagnose and prediction architecture, the PHM technique can find its applications from the equipment level to the platform level, which provides the aircraft an all-round health management.

Specifically, the on-board PHM system involves three levels. The bottom level includes these monitoring software and hardware distributed in the devices of the aircraft, such as the sensors and the built-in test (BIT) equipment. The middle level

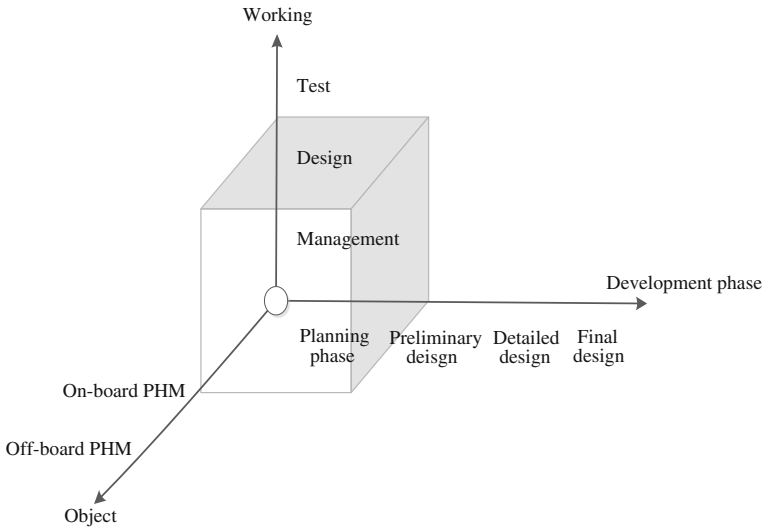


Fig. 1 A three-dimension view of the PHM development process

includes the regional managers and the top level is the aircraft platform manager. The bottom level serves as the fault information source: it collects the fault information via the sensors and the BIT and then submits the information to the regional manager in the middle level. The regional managers have the capacities of signal processing, intelligence fusion and regional reasoning, which monitor the corresponding subsystems of the aircraft continuously. The aircraft level PHM manager locates in the Integrated Core Processor (ICP). By applying the association rules, it can synthesize different information, identify and isolate the fault, and eventually formulate the maintenance information and provide available knowledge for the pilot. The maintenance information would be transmitted to the Automatic Logistics Information System (ALIS) on the ground. Then ALIS will synthesize and identify the information provided by the on-board state manager and make decisions based on the information.

The working dimension reflects the works with different natures involved in the PHM development, including the design process, the management of the development and the validation test of the system. Specifically, the PHM design includes all the design works completed in different development phases. The management of the PHM development involves making the PHM design working plan, decomposing and distributing the overall requirements, managing the design documents and drafts during the design process and periodically evaluating the design. The test acts as the validation for the PHM system after the design is completed; it includes the systematical simulation validation and tests of the PHM system and its subsystems.

In the development dimension lies phases of the PHM development procedure: the planning phase, the preliminary design, the detailed design and the final design. This dimension is a time line of the PHM development and the decomposition of the PHM system is implemented along this dimension.

3 WBS of the PHM System

With the discussion in Sect. 2, the different aspects of the PHM development become clear in three dimensions. In this section, we will decompose the PHM development in this three-dimension view with the WBS technique. Specifically, the PHM development will be decomposed in the object dimension and in the working dimension. The design in the working dimension is discussed in Sect. 4 in detail and hence will not be discussed in this section.

The WBS is defined as a hierarchical description of the involved works during the design and manufacturing of the equipment by a top-down stepwise decomposition [1]. The hierarchical description focuses on the product that will be designed and manufactured. The works in the equipment project and their relations with the final product is completely defined by three elements, i.e., the product, the service and the documents. The characteristic of the WBS is that it represents the procedure of decomposing a complex project sequentially into some simple elements [6].

3.1 Decomposition of the Object (the PHM System)

Along the object dimension, the PHM system can be decomposed into four levels according to the product function by a top-down manner, which is shown in Fig. 2. For the fourth level, we summarize different parts with functional relations and obtain six modules: data acquisition module, data manipulation module, state detection module, health assessment module, prognostic assessment module and advisory generation module. Combined with the interfaces, including the man-machine interface, interfaces between modules and interface with other systems, it arrives at a typical PHM structure of open system architecture for condition-based maintenance (OSA-CBM).

3.2 Decomposition of the Works: Management

In the working dimension, the management involved in the PHM development is decomposed into the systems engineering management and the engineering project management, as shown in Fig. 3.

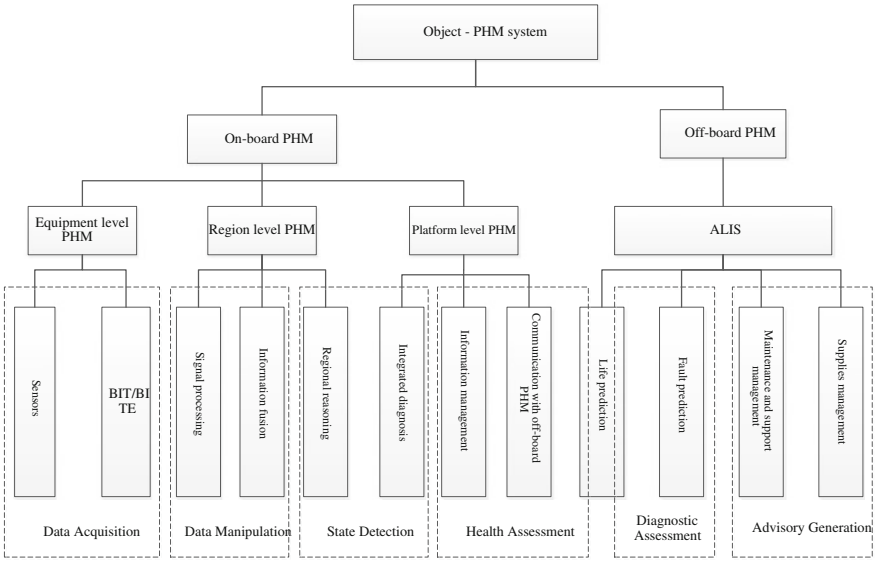


Fig. 2 Decomposition along the object dimension

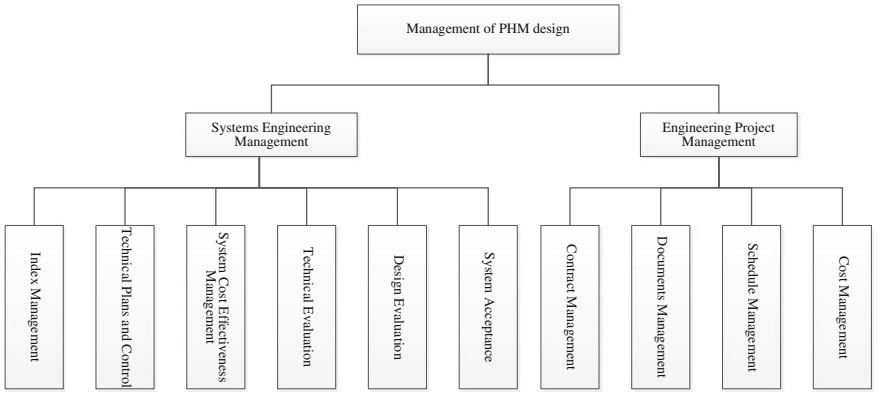


Fig. 3 Decomposition along the working dimension: management

The systems engineering management is a complex task. It includes six elementary management works. The target of the index management is to propose the monitoring indexes according to the requirements on the PHM system, such as the fault detection rate (FDR), fault isolation rate (FIR) and false alarm rate (FAR). These indexes will be decomposed into the more detailed indexes according to the characteristics of different subsystems and their working conditions. The detailed indexes will be distributed to the subsystems and will be their design requirements. Technical plans and control deals with the technical planning documents and

technical controlling documents in supervising, guiding and accessing the technical outline, as well as the standard system that has to be clarified in the PHM design process. System cost-effectiveness management includes the risk analysis and the cost-effectiveness analysis of the PHM system, considering that the introduction of the PHM system also increases the deployment cost and the system risk while it can predict faults and reduce the support cost. Technical evaluation is to review the qualification of the specialized designing departments before assigning the task to them. Design evaluation is to check the design of the PHM system periodically and identify whether the requirements are met and the process is carried out according to the schedule. System acceptance includes compiling the test plan and validating and identifying the PHM system.

The engineering project management is also a complex task. It focuses on these issues that involved in the PHM development but excluded by the systems engineering management, including the contract management, documents and drafts management, schedule management and cost management.

3.3 Decomposition of the Works: Test

The test in the working dimension is a generalized concept, which includes the simulation test and the real PHM system test. The aim of the tests is to validate and access the PHM system and modify the possible design defects. Figure 4 illustrates the decomposition of the test for the PHM system [9, 10].

The simulation test is applied to the systems that subject to high safety requirements and not suitable for real system tests. These systems are modeled and

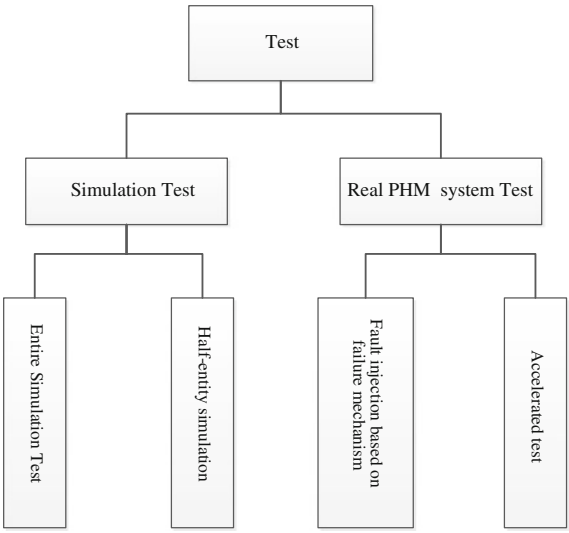


Fig. 4 Decomposition along the working dimension: test

simulated to validate its performance and the models will be modified when the historical data are available to improve the accuracy of the fault location. Generally, the simulation test will cost less than the real system test and can reduce the number of required real system tests. On the other hand, the simulation test cannot perfectly reflect the environmental impacts on the system and requires lots of efforts on the system modeling and simulation. The test on the real PHM system can be implemented resorting to the fault injection technique or the accelerated test. The system performance can be evaluated with the operations of the target system or its equivalent system.

The design in the working dimension involves a sequential PHM design process based on the characteristics of the PHM system and the aircraft and their working conditions. The PHM design process includes requirement analysis, requirement decomposition, design and implementation, function integration, analysis and evaluation, maturation and validation and identification. The details of the PHM design will be discussed in Sect. 4.

4 Collaborative PHM Design Process

The structure of the PHM system and the involved works becomes clear after decomposing the PHM system in the object dimension and in the working dimension. In this section, we will further discuss the PHM design in a collaborative view. Indeed, all the works, including management, design and test, play their roles though the whole PHM development cycle; they should be carried out simultaneously in a collaborative way.

For the aircraft, its design involves multiple-disciplines and requires engineers from different domains or different places to cooperate in a same project. For the PHM system, its design requires the design information about the critical systems of the aircraft, which demands the cooperative works of the aircraft designer and the PHM designer to achieve an optimal design. Considering that the collaborative design is distributed, interactive, sharing, dynamic and collaborative, which is consistent with the requirements of the aircraft PHM design, it is reasonable to introduce the method of collaborative design into the aircraft PHM design.

The collaborative design carries forward the idea of concurrent engineering, where a group collaboratively completes a design task under a computer supported environment. It is a new design technique that requires the designers considering not only the technical issues involved in the design of the product but also the user's requirements, the manufacture process, the assembly and the maintenance. The core concept of the collaborative design is the optimization of the system and the integration of the development, anticipating all the problems that might arise in the development and proposing the corresponding solutions at the very beginning of the design process [7]. Under a collaborative design environment, the PHM design cycle will be shortened and the PHM system can be rapidly developed. Designers and experts can communicate on the collaborative working platform and

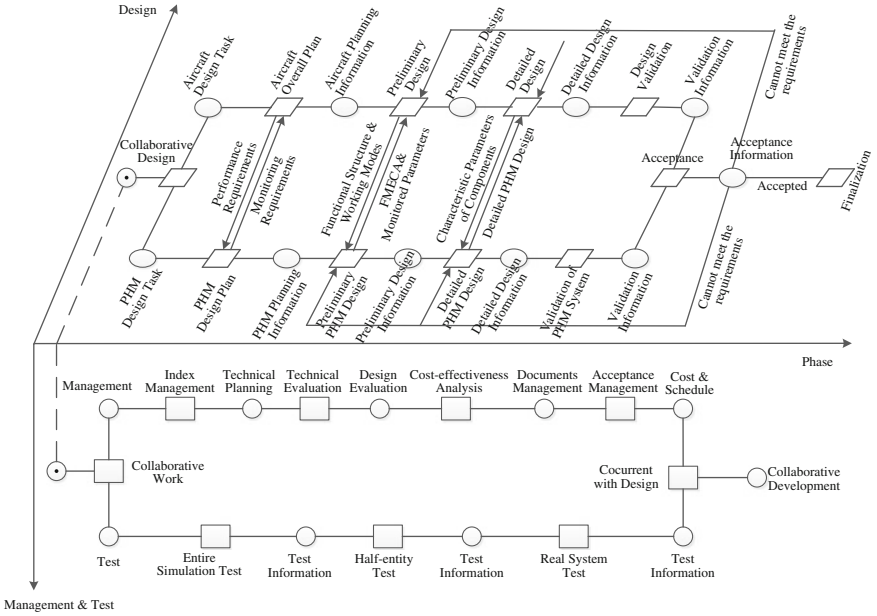


Fig. 5 Collaborative PHM design framework

solve the problems more efficiently. The design process can achieve a high level integration with a seamless data transition and model simulations. The development cost can be estimated more accurately and reduced by integrating the cost model and the cost design process [5].

Figure 5 illustrates the framework of the collaborative PHM design. There are two levels of collaboration in the PHM development: for the PHM system, the involved works including management, design and test are carried out collaboratively; for the PHM and the aircraft, their design processes proceed collaboratively. In the followings, we will discuss these two levels of collaboration in detail.

4.1 Collaborative PHM Design and Aircraft Design

The aircraft design consists of the concept demonstration phase, preliminary design phase, detailed design phase and the finalizing design phase. Similar, the design process for the PHM system also contains these four phases. The specific design variables of the PHM system are defined by the design of the components, subsystems and systems of the aircraft, that is, the PHM design has to accommodate the aircraft design. In order to achieve an optimal overall design, the PHM design and the aircraft design should be carried out simultaneously and collaboratively.

In the planning phase, the tactical requirements and the overall concept solution of the aircraft can be settled according to the users' demands and their feasibility will be validated. Accordingly, with the monitoring requirements of the aircraft, the overall PHM solution will be proposed and be accessed to see whether it can meet the monitoring requirements.

In the preliminary design phase, the technical solution of the aircraft design, i.e., the layout and the overall design parameters will be proposed according to the development requirements. At the same time, the PHM design elements will be decomposed according to the structure, function and working modes of the systems in the aircraft and incorporated into the aircraft design. Specifically, the structure of the PHM monitored object should be identified and decomposed; then the failure modes, mechanism, effects and characteristic parameters of the object should be identified resorting to the FMECA technique. Whenever the weak parts of the monitored object are identified, the reasoning framework of fault diagnosis and prediction can be constructed.

In the detailed design phase, the design will be refined according to the preliminary design plan, and the characteristic parameters of different parts of the aircraft, such as the shape, the size, the materials and the manufacturing, will be confirmed. Prototypes will also be built to verify the design. For the PHM system, the sensors locations are defined according to the characteristics of the monitored object. The adaptability of the fault diagnosis algorithm will be analysed and the design of the reasoning machine will be refined. Proper prediction algorithms should be selected to predict the health status of the components, systems and the aircraft, and the maintenance decision and support plan generation system should be designed. Maintenance decisions are mainly generated by the off-board PHM equipment, whose design includes the hardware and software design and the interfaces design, such as the interfaces between modules, interfaces between different levels, man-machine interfaces and the interface between PHM system and the aircraft.

In the finalizing design, after the prototype is built, ground tests such as static mechanical test, resonance test and electromagnetic compatibility test have to be implemented before the test flight. Whenever the failure occurs during the test flight, it has to be investigated and understood, and modifications have to be done whenever necessary. Both the design of the aircraft and its PHM system are validated in this phase. The validation test of the PHM system mainly resorts to the half-entity test and the real PHM system test.

A joint test of the aircraft and the PHM system should be accomplished after the two systems are validated separately. The aim of the joint test is to verify the compatibility of the PHM system and the aircraft and modify the design whenever the requirements are not satisfied. The process is iterative until all the design requirements are met.

4.2 Collaborative Management and Test

Management and test verification continue playing their roles throughout the whole PHM development process. They have to be carried out collaboratively with considering the ongoing design process.

For the management works during the PHM development, with the decomposition in Sect 3.2, the works to be carried out during each design phases become clear. During the planning phase, the management works include the indexes and requirements management and the technical planning. The standard system and the technical guidance outline of the PHM design should be proposed, based on which the following design works can be carried out. Besides, other works including the cost-effectiveness analysis, the risk analysis, the technical evaluation and the task distribution for the specialized designing departments should be carried out at the same time. In the design phase, i.e. the preliminary and detailed design phase, the management works includes the periodical design evaluation and the schedule controlling. The related documents should be filed and will form the knowledge base for future projects. In the finalizing phase, the management works involve verification of the project and the final acceptance, including the design acceptance and the cost acceptance.

For the test works involved in the PHM development, the work elements involved in each design phase can also be identified with the decomposition in Sect. 3.3. During the planning phase, tests to validate the feasibility of the new techniques should be carried out. In the design phase (the preliminary design and the detailed design), the major part of the test works is the entire system simulation, by which the design of the system can be validated and modified. In the finalizing phase, more tests are carried out on the real PHM system by fault injection technique to verify whether the system can locate the fault properly and give admirable maintenance suggestion. The accelerated test, such as the accelerated lifetime test or accelerated degradation test can also be applied to access the safety and the reliability of the PHM system. At the same time, the joint tests on the PHM system and the aircraft have to be carried out to verify the compatibility between the PHM system and the aircraft.

5 Conclusion

Currently, the aircraft design and its PHM system design are carried out separately and as a result the overall optimal design can never be reached. This paper introduced the collaborative design method into the PHM system design. First, the PHM development is analyzed in a three-dimension view: the object (PHM system) dimension, the working dimension and the phase dimension. Then the PHM development process is decomposed in the object dimension and in the working dimension by applying the WBS technique. After the decomposition, the structure

of the PHM development process becomes much clearer. Then we proposed a collaborative PHM development framework based on the collaborative design method, in which the PHM design and the aircraft design are integrated collaboratively and the management, the design and the test in the PHM development are also carried out collaboratively. The benefits of the collaborative PHM development lie in that it cannot only optimize the overall aircraft design but also will reduce the development cycle and the lifetime cost. In the future, we will provide a software platform based on the method provided in this paper to aid the PHM development process.

References

1. GJB 2116–1994 Weapon Equipment Development Project Work Breakdown Structure
2. Hess A (2002) Prognostics, from the need to reality-from the fleet users and PHM system designer/developers perspectives. IEEE aerospace conference proceedings, vol 6, pp 6-2791
3. Hess A, Fila L (2002) The joint strike fighter (JSF) PHM concept: potential impact on aging aircraft problems. IEEE aerospace conference proceedings, vol 6, pp 6-3021
4. Hess A, Calvello G, Dabney T (2004) PHM a key enabler for the JSF autonomic logistics support concept. IEEE aerospace conference proceedings, vol 6, pp 3543–3550
5. Monell DW, Piland WM (2000) Aerospace systems design in NASA's collaborative engineering environment. Acta Astronaut 47(2):255–264
6. Pi YF (2006) Study on spaceflight model work breakdown structure (WBS). J North China Inst Aerosp Eng 16(3):1–3, 8
7. Rui YN (2003) Cooperative design. China Machine Press, Beijing
8. Sun B, Zeng S, Kang R, Pecht M (2010). Benefits analysis of prognostics in systems. IEEE prognostics and health management conference, pp 1–8
9. Xu P, Wang ZL, Li V (2010) Prognostics and health management (PHM) system requirements and validation. IEEE prognostics and health management conference, pp 1–4
10. Yang Z, Jing B, Zhang J, An YJ (2012) Verification and evaluation method of airborne PHM system. Meas Control Technol 31(3):101–104, 111
11. Zeng SK, Pecht MG, Wu J (2005) Status and perspectives of prognostics and health management technologies. Acta Aeronaut ET Astronaut Sin 26(5):626–632

Managing Knowledge Assets for the Development of the Renewable Energy Industry

Chung-Shou Liao, Hung-Yu Huang, Sheng-Ting Yang and Amy J.C. Trappey

Abstract While the world is increasingly concerned with the utilization of fossil energy and carbon emission, renewable energies have been promoted by many governments as alternatives worldwide. Currently, many databases provide big energy data such as international crude oil prices, coal prices, etc., but lack systematic analysis. To increase the use of the data, we propose GEIP (Green Energy Information Platform) using data retrieval techniques to develop a knowledge management platform which can integrate inconsistent information and analyze the development of renewable energy industries. GEIP provides an easy-to-use interface as well as a comprehensive query function from multiple data sources. In addition, the proposed platform can also simultaneously consider distinct factors influencing the development trends of renewable energies, when the development of the renewable energy industry is actually determined by multiple related factors. The result demonstrates our system well project to the development trends of wind power and solar PV in several countries.

1 Introduction

In the last decade, due to the increasing focus on environmental issues, Taiwanese government has aggressively promoted renewable energy. There are two reasons. First, Taiwan is both a petroleum- and coal-dependent country, and the proportion

C.-S. Liao (✉) · H.-Y. Huang · S.-T. Yang · A.J.C. Trappey
Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan
e-mail: csliao@ie.nthu.edu.tw

H.-Y. Huang
e-mail: s100034518@ie.nthu.edu.tw

S.-T. Yang
e-mail: s10134554@ie.nthu.edu.tw

A.J.C. Trappey
e-mail: trappey@ie.nthu.edu.tw

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_143

and coal supply is about 80 % every year [2]. Second, the consumption of traditional fossil fuels obviously causes greenhouse gas emissions, and Carbon dioxide emission per capita in Taiwan is getting much higher than the world average [6]. Thus, Taiwanese government has approved the sustainable energy policy program in order to enhance the energy efficiency, increase the green energy utilization, and assure the stable power supply of the country.

To promote the development of renewable energy, the governments have implemented sustainable energy policies and enhance the knowledge management of the latest energy information. On the other hand, advances in information technology have much improved systems-level information management. Currently, many researchers start to employ information technology for knowledge management in different areas such as energy management. Although there are many up-to-date databases comprised of energy information as well as the corresponding web services, the collected data may not consistent with each other. Their interfaces are also difficult to access for the public. Therefore, it is worthwhile to filter the collected energy data and effectively provide numerous scientific decision supports.

In this study, we developed GEIP which is an information management system as well as a decision-support analysis service platform—GEIP aims for exploiting data retrieval techniques to establish a knowledge management platform to present real time energy information and analyze the development of renewable industries. In particular, our management platform can be dynamically connected with the related well-known databases and also bound with different kinds of information systems. We conducted experiments for wind power and solar PV, which are currently the most popular renewable energy in Taiwan as case studies. We demonstrate the result through a variety of interactive user interfaces for different communities, such as the government and private sectors, academic researchers, and the general public. We believe that the proposed platform can effectively promote the communication of energy information.

2 Knowledge Management Platform on Energy Issues

In recent years, some research platforms have been constructed via service interfaces (see Table 1). Most of them mainly focused on the areas of medical science and health care. For instance, Teijeiroa et al. [5] built a distributed platform for the home supervision of multiple diseases. Pablo et al. presented the design of a mobile gateway for independent life and e-health support [4]. Accordingly, this study attempted to apply the similar concept to conducting an information service platform in energy area. Table 1 shows the comparative analysis between other systems-level platforms and the proposed platform.

There are three stages in GEIP: real time data collection, the platform construction, and decision support system for assessment. The first stage, data collection, is to mine and filter related well-known databases around the world, based

Table 1 Comparative analysis of other system platform and our platform

Systems platform	The United Nations Statistics Division (UNSD)	Centers for Disease Control and Prevention (CDC)	National Statistics, Taiwan	Green Energy Information Platform (GEIP)
Highlights	The platform compiles and disseminates global statistical information in social, energy, economic and engineering area, and develops standards and norms for statistical activities. They also support countries' efforts to strengthen their national statistical systems	The platform provides users with credible and reliable health information on diseases, conditions, and environmental health	The system provides economic, social and environmental information of Taiwan for different users (public, students and academia)	This platform employs the information retrieval techniques and dynamically connected with the related databases. It provides economic and energy information as well as decision-support analysis for public and academia

on the interviews with the experts in Green Energy and Environment Research Laboratories of the Industrial Technology Research Institute (ITRI). We employ the techniques of Database Management Systems (DBMS) to connect, manage and monitor all collected data in the second stage. This study has built a knowledge management platform on energy issues. In addition, we continuously evaluated and modified the capability of this platform.

In particular, our platform provides an analytical decision-support service for the development of renewable energy. This service employs the hidden Markov model (HMM) [1] to determine the development of renewable energy via multiple influential factors. We tested the supporting service on the industry of wind power and solar PV in Taiwan as case studies. The analysis result can be used to assist the industry in making correct investment decision and to provide recommendations to these governments for setting suitable green energy policies.

3 Contents and Features

We briefly discuss the features of GEIP which presents user-friendly interfaces. There are three main features in GEIP: information integration and management, comparative view and smart analysis. We introduce main purposes and procedures as follows.

Information Integration and Management: GEIP provides energy-related information browser functions. GEIP integrated several well-known databases so that users can easily get the real time data from the selected database by simply checking on/off the checkbox. The procedure contains two major steps (Fig. 1a). The first is to choose the options that meet your needs, such as "Data source",

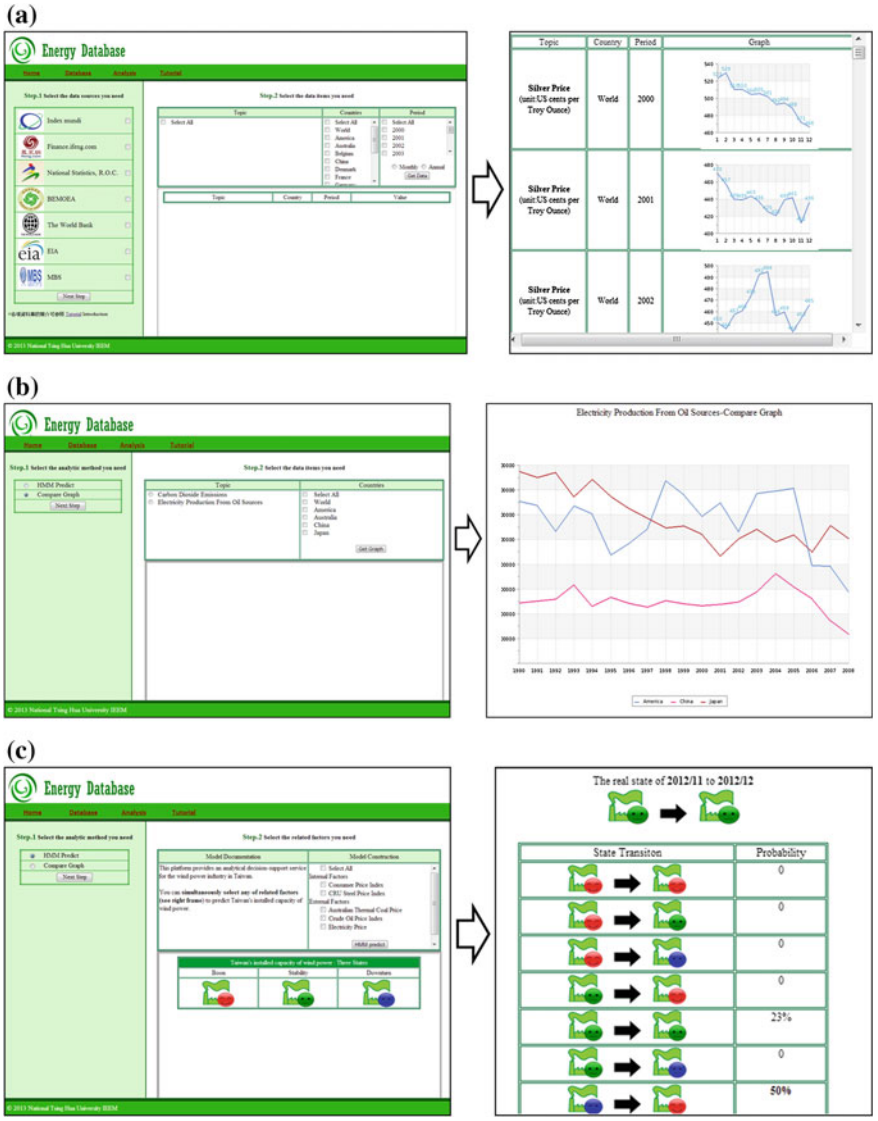


Fig. 1 **a** The information browser feature. **b** The graphical comparison feature. **c** The smart analysis procedure

"Topic", "Countries" and "Period". Next, users press the "Get data" button, and then our system will automatically compile the information and draw line plots out using the JpGraph tool [3]. This feature simplifies the information retrieval step, and reduces the interface diversity of different databases and web tools.

Comparative View: Compared with other tools and databases in the literature, GEIP provides the capabilities of data processing and graphic support. User can conduct a preliminary analysis and comparison through graphic features. Two operations are described as follows (Fig. 1b). The first step is to choose the "Topic" and "Countries" according to your requests. Next, users can press the "Get graph" button, and the system will automatically draw figures to graphically represent the queried data. The system can support many graphic formats such as line plots, bar charts, pie plots and so on using the JpGraph tool.

Smart Analysis: In contrast with most of other query tools and databases, GEIP contains an intelligent analytical decision-support service for forecasting the development of renewable energy industries. Two steps for this function are as follows (Fig. 1c). First is to select HMM predict and "Model Construction" which you consider. Then, the system will show the prediction result associated with probabilities. This feature can be used to assist the industry in making correct investment decision and to provide recommendations to governments for setting suitable green energy policies. It would be helpful to the emerging renewable energy industries.

4 Conclusion

The energy issues have become the focus of national policies and public due to the rapid energy shortages. Therefore, in order to promote energy knowledge management and the development of renewable energy, this study employs data retrieval techniques to develop a knowledge management platform. The platform also provides decision-support interface using HMM to analyze the development of wind power and solar PV in Taiwan. The result demonstrates that the information service of our systems-level platform has been presented for different users. It would be of great interest if there could be more analytical decision supporting models included in the near future.

References

1. Baum LE, Petrie T (1996) Statistical inference for probabilistic functions of finite state markov chains. Ann Math Stat 37(6):1554–1563
2. Bureau of Energy, Ministry of Economic Affairs. Energy Statistical Data Book (NO. 2011). Retrieved from http://web3.moeaboe.gov.tw/ECW/populace/content/wHandMenuFile.ashx?menu_id=144
3. JpGraph, A PHP-driven Charts Library. Retrieved from http://jpgraph.net/

4. García-Sánchez P, González J, Mora AM, Prieto A (2013) Deploying intelligent e-health services in a mobile gateway. Expert Syst Appl 40(4):1231–1239
5. Teijeiro T, Félix P, Presedo J, Zamarrón C (2013) An open platform for the protocolization of home medical supervision. Expert Syst Appl 40(7):2607–2614
6. The World Bank. Retrieved from http://data.worldbank.org/indicator/EN.ATM.CO2E.PC/countries/1W?display=graph

Dynamic Patent Analysis of Wind Power Systems and Engineering Asset Development

Amy J.C. Trappey, Chii-Ruey Lin, Chun-Yi Wu and P.S. Fang

Abstract Patent analysis of new technology development can be focused on various aspects, e.g., identifying the competitiveness of a specific country or the technical advances of a firm in the given domain. The offshore wind power (OWP) has been promoted and encouraged by many countries for their renewable energy development. This research focuses on systematically discover the profile of OWP technology innovation extracted, analysed and synthesized from world patent databases dynamically using text and data mining techniques. In real time, all related patents are searched and collected. Afterward, the research builds an OWP ontology schema by identifying the OWP technical domain structure, the core concepts, and the relationships between the concepts. The ontology schema is constructed and refined dynamically using key phrases automatically extracted from patents and verified by domain experts. The research also emphasizes on the control of risk and reliable factors when evaluating the offshore wind power development. Thus, the research applies patent indicators, i.e., the patent family, the growth rates of domain technologies, the reference and citation network, the technology and function matrix, and the patent litigation information, to derive the OWP R&D landscape and strategy. After building the OWP patent knowledge base (KB), the potentials of OWP development opportunities for a specific region can be clearly identified with reliable measures based on patent map informatics.

A.J.C. Trappey · C.-Y. Wu (✉) · P.S. Fang
Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan
e-mail: d9534524@oz.nthu.edu.tw

A.J.C. Trappey
e-mail: trappey@ie.nthu.edu.tw

P.S. Fang
e-mail: s101034510@m101.nthu.edu.tw

C.-R. Lin
Department of Mechanical Engineering, National Taipei University of Technology, Taipei, Taiwan
e-mail: crlin@ntut.edu.tw

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_144

Keywords Offshore wind power · Ontology · Patent strategy · IP risk management · Critical technology of green energy

1 Introduction

Offshore wind power depends on wind-driven generator to produce electricity. Wind power generator array forms an offshore wind power farm on water, such as sea or lake. In Taiwan, the best place for developing offshore wind power is western Taiwan Strait. The benefits of offshore wind power can be divided into both environmental and economic aspects. In the environmental aspect, the development of wind power, substituting fossil fueled energy, reduces carbon dioxide (CO2) emissions significantly. In the economic aspect, the development of offshore wind power promotes the related industries. Manufacturers of equipment and marine construction can be organized as the OWP supply chain. The advantages of developing offshore wind power, instead of onshore wind power, are as follows. First, offshore wind power speed is usually faster than onshore by 20 % and its grabbing wind energy increases 72 %. Second, the offshore airflow is more stable than onshore, the wind power generator fatigue load is smaller, and the life of offshore wind power generator is longer than onshore by 25 %. Third, offshore wind power farm is away from land where problems of noise and lighting are fewer than onshore. Thus, it is plausible to increase wind speed and efficiency. Forth, annual full load hour is longer than onshore and there are advantages to expand generating capacity. Finally, it is easier to achieve economic scale and shorten the recovery period. Offshore wind power generator will develop toward the goal of lower generating cost and increasing unit capacity.

Taiwan, with special geographical conditions has high potential of natural disasters, such as typhoons and earthquakes. Although the 20 m depth offshore wind power farm is a matured development, it is difficult to use directly in Taiwan. We have focused on research of onshore wind power farm and key components of wind power generators in the past. Therefore, OWP technology development and industries have not started in full scale. Taiwan energy sector hopes to use the existing wind power technology in combination with new technology to support OWP key industries. In the past, the static patent analysis identified interesting patents in the one-off analysis. In this research, the procedures of dynamic patent analysis use automatically collected key phrases from patent text mining to build and update OWP ontology. Through routinely searching patent databases, the ontology is renewed constantly with the latest data. Therefore, the critical and newest patents of OWP technology are incorporated into the ontology-based landscape. This research focuses on analyzing and synthesizing OWP technology innovations extracted from USPTO (USA), EPO (Europe) and TIPO (Taiwan) patent databases. Thus, the OWP development strategies and intellectual property (IP) opportunities and risks can be better assessed and managed for the policy makers of green energy and the domestic OWP industries.

2 Literature Review

This section reviews the relative research literatures, including offshore wind power technology, ontology, patent map analysis, and the definitions and purposes of growth curve.

In this research, we focus on the domain of offshore wind power system. A simplest wind-energy turbine consists of three crucial parts. First, the blades are the basic sails of the system. In their simplest form, they act as barriers to the wind (more modern blade designs go beyond the simple barrier concept). When the wind forces the blades to move, it has transferred some of its energy to the rotor. The wind-turbine shaft is connected to the center of the rotor. The rotor spins will cause the shaft to spin. The rotor transfers its mechanical and rotational energy to the shaft, which enters an electrical generator on the other end. A generator uses the properties of electromagnetic induction to produce electrical voltage. Voltage is essentially electrical pressure—it is the force that moves electricity or electrical current from one point to another. The voltage drives electrical current (typically alternating current, or AC power) through power lines for distribution [10]. Based on the structure of the OWP system, the research uses an ontology to represent the concepts and relationship of the domain knowledge for automatic patent search. Ontology defines common vocabularies for ones who need to share information in a domain of interest [1]. Ontology also provides general criteria for the segmentation and the organization of domain [9]. It includes machine-interpretable definitions of basic concepts in the domain and relations among them [15]. Both domain-dependent and general ontology need to be presented in formal structures, which can be understood and shared by many parties.

After collecting the structure of domain technology and building domain ontology, the research collects the domain patents from published patent databases to analyze the development of domain technology. Patent analyses have often been employed as economic indicators that measure the linkages between technology development and economic growth [4]. Recently, the strategic importance of patent analysis is highlighted in high-technology sectors as the processes of innovation becomes complex, the cycle time of innovative R&D is shorten, and the market demands are hard to predict. The practical value of patent analysis is high among technology-oriented companies. They use patent analysis to estimate technological knowledge trends, impacting on market advantages and profitability, and comparing innovative performances in the international context [7, 12]. As Granstrand [3] described many types of patent deployment models, e.g., blocking, designing/inventing around, strategic patenting, blanketing, flooding, fencing, surrounding, and combination of strategies. Successful patent analyses not only save cost of enterprise R&D, but also strengthen R&D efforts offensively and defensively. Patent analysis help to interpret the status and trends of technology development by utilizing patent bibliometric method, which is quantitative oriented to search out the key technologies in related industries [6]. Moreover, a patent can be categorized into specific function and the methods of function-based patent analysis identify

scopes and potential application industries of technology by evaluating the important patent in each industry [8]. For example, the approach to analyze technology life cycle (TLC) is mainly S-curve by adopting multiple patent-related indicators to help understand technological performance, such as citation, application, assignee, inventor, and so on [2]. Therefore, this research uses various data mining approaches to rapidly and automatically analyze the features of patents and extract the characteristics of patents. Furthermore, the research also uses the patent growth curve model to evaluate the level of the technology development. The definitions of growth curve include Upper Asymptote, Point of Inflexion, and Lower Asymptote. Upper Asymptote represents the upper bound of growth curve as the technology is well-developed. Point of Inflexion describes that the technology is at the fastest growth rate. Lower Asymptote is defined as the lower bound of growth curve as the technology being at its early growing stage.

3 Methodology

In this research, we deploy both quantitative and qualitative analyses of the OWP patent information for evaluating the trend of offshore wind power technology innovations. For quantitative analysis, we analyze the changes in the total number of patent applications over the years from different patent databases (USPTO, EPO, and TIPO). The changes in the numbers of applications and the numbers of issued patent are divided based on different technical fields (IPC and UPC). Afterwards, this research analyzes the situation in the percentages changes of patent applications by technical fields from different assignees and countries. Moreover, the patent technologies and functions are depicted in a matrix to identify the relationship of potential innovative technologies. We also evaluate the forward citation ratio and its patent family of patent documents to rank the importance of patents in domain technology. In the qualitative analysis, this research uses the text mining to extract the key phrases of domain important patents from the content of patent abstracts, claims and detailed descriptions. The extracted key phrases help build the domain ontology covering different technology sub-fields. The combined quantitative and qualitative analyses provide enterprises with versatile decision support dashboard and views in finding the opportunities of technology development.

3.1 Patent Map Analysis

In a qualitative analysis, the research extracts the important patents and analyze diagram of development and changes under multidimensional classification. After that, we can find the changes in the number of examined publications under multidimensional classification. Moreover, the research calculate three-structured diagram of development and diagram of effects of technologies introduced or spread

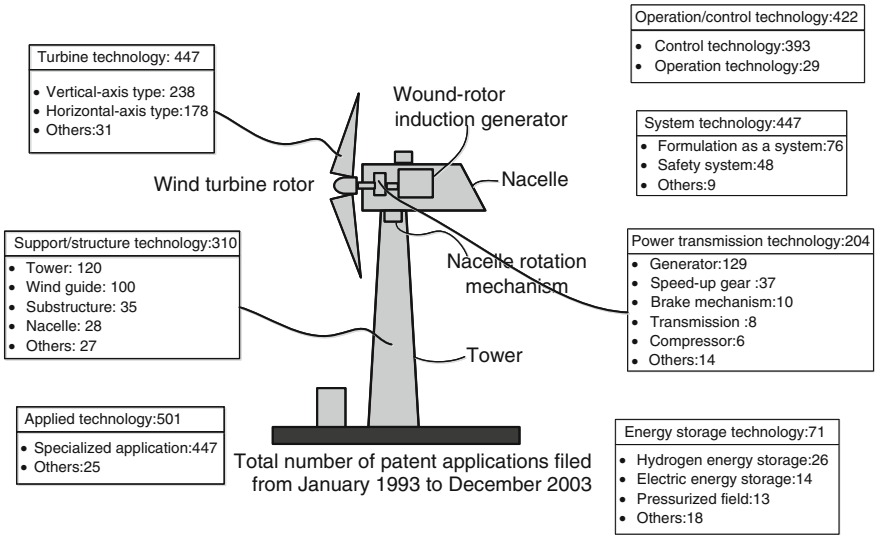


Fig. 1 Illustration of wind turbine generator and patent statistics in categories [13]

from other technical fields. Thus, we can evaluate overall diagram of the association of technological development. One of the most basic Patent Maps is to collect patent documents for a particular right-holder, arrange them by year of filing of patent application, and plot the number of patents or patent applications. This is called a Time Series Map. A Time Series Map is used to analyze the trends and numbers of patent applications filed by or patents issued to specific assignees. Figure 1 shows a systematic art diagram of wind-turbine generator based on its technical elements and the numbers of patents granted according to the elements [13]. A wind-turbine generator involves: (i) blade technology that is used to convert wind power into rotational kinetic energy; (ii) power transmission technology that is used to transmit the rotation of turbines to a power generator; (iii) support/structure technology; (iv) operation control technology; (v) system technology; (vi) energy storage technology; and (vii) applied technology.

A qualitative analysis, including a systematic art diagram, must meet the requirement that the patent collections should be complete and that the irrelevant patents ("noise") should not be included. Specifically, given the differences in the classification schemes between industrial nomenclatures and patent classification codes (IPC, UPC), in order to collect relevant patents without omission, it is important to visually check the basic data by the industrial specialists. If an important patent is missing, it should be retrieved and included in the collection.

3.2 Assumptions of Growth Curve

For assumptions of growth curve [5], the research needs to follow the conditions. First, upper bound of growth curve is known. Second, the selected growth curve meets the change of historical data. Third, the coefficient of the growth curve formula presents a good match of the historical data. Pearl curve is known as logistic curve, where y is a non-linear function of t. L is the upper limit of y. This method has been applied for population forecasting. Pearl curve is shown in Formula (1).

$$y = \frac{L}{1 + \alpha e^{-bt}} \tag{1}$$

y = Measuring technology effects (i.e., the cumulative patent counts).
L = Upper limit to the growth of the variable y (i.e., the maximum of cumulative patent counts).
e = Base of the natural logarithms.
t = Time (i.e., the patent application year).
a, b = Coefficients obtained by fitting the curve to the data.

The properties of Pearl curve are as follows.

- Initial value of zero at time = $-\infty$ and a value of L at time = $+\infty$.
- If the initial value is not zero, the initial value can be added as a constant to the above formula.
- The inflection point occurs at t = ln(a)/b, when y = L/2.
- The curve is symmetrical at the inflection point, with the upper half being a reflection of the lower half.

3.3 Wind Power Ontology

After using patent map analysis and building the growth curve, the research collect the domain patent to develop the ontology of offshore wind power. The research searches the OWP patents from USPTO. The key phrases of collected patent are extracted by text mining, which is to separate the sentences by using Term Frequency-inverse Document Frequency (TF-IDF) approach. Salton and Buckley [11] integrated the TF and IDF to create the TF-IDF value. The TF-IDF checks to see if the appearance of a phrase is higher in one document and lower in another document. Thus, the research extracts the key phrases, including blades, rotor, pitch, brake, gear box, generator etc., as shown in Fig. 2. For the automatically extracted key phrases, the synonym counts are consolidated using the patent terminology dictionary [14]. The system collects other domain patents from EPO and WIPO

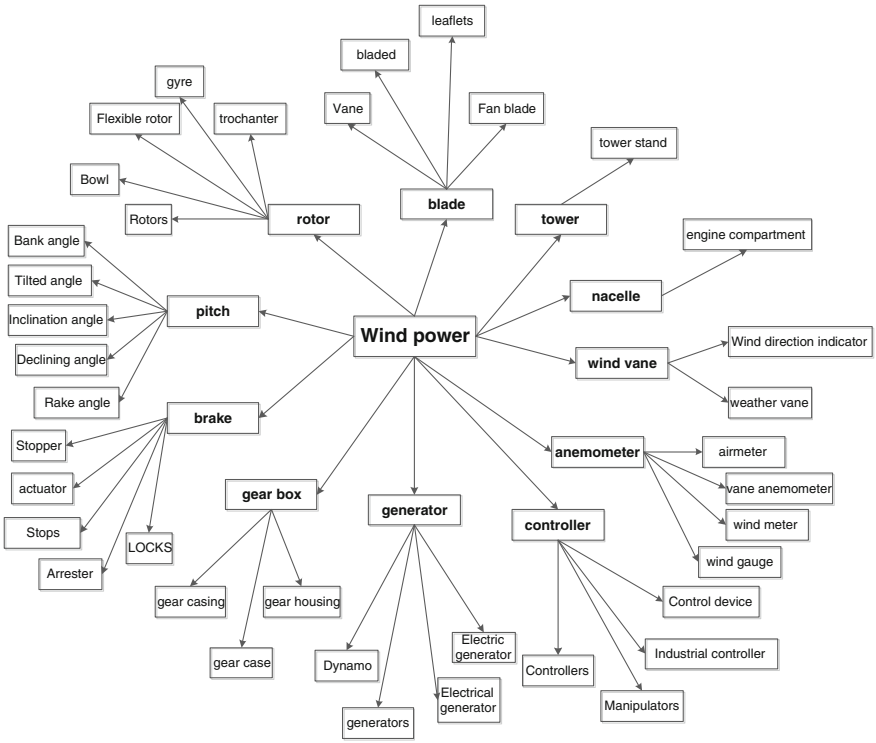


Fig. 2 The key phrases and synonym of the wind power technology

patent databases after USPTA and TIPO database search. All the OWP patents are constantly updated, classified, and analyzed on a system platform. The experts from academic and industry provide the structure of wind turbine system and help to confirm the accuracy of key phrases and ontology. Finally, the OWP ontology is constantly maintained and updated to help IP engineers search other related patents from global patent databases.

4 Case Analysis of OWP Patents

This research intends to understand the development life cycles of wind power technologies. Through United States patent and trademark office (USPTO), we consider application scope of wind power generator components from key phrases and IPC. First, we understand key phrases of this project by technology terms and use main key phrases, such as wind power or wind energy, and secondary key phrases, such as blades, rotor, generator and tower. We use the key phrases appeared in patent titles and abstracts as search conditions to find relevant patents

Table 1 An example of patent search condition

Search column	Title, abstract, claim, application date
Key phrase	Wind, generator, blades, rotor, tower
Database name	USPTO
Search year	To 12/31/2012
Analysis date	Patent issued date
Patent search condition	((TTL/"wind" and ABST/"generator") and ACLM/"blades")
	((TTL/"wind" and ABST/"generator") and ACLM/"rotor")
	(((TTL/"wind" and ABST/"generator") and ACLM/"generator") AND APD/20050101→ 20121231)
	((TTL/"wind" and ABST/"generator") and ACLM/"tower")
Patent number	1443
Final result	750

and 4,549 patents are found. We keep on expanding search constraints by adding "generator" to search patents and get 877 patents. By using "generator" and "blades" as search conditions, we narrow the search results to 304 patents. Finally, we use "wind power" in title, "generator" in abstract and "wind turbine technology" terms in claim, 482 patents were identified (e.g., Table 1).

From Fig. 3, we can understand technology of wind power generator that mainly applies for patents in 1975–2012. It's the initial development stage in 1975–1981 because of energy crisis. Due to the international oil price stability, it's a stable development stage in 1982–1999 and wind power generator research is in the stagnation period. There seemed be slow development from 1975 to 2000 in wind power technology. Patent applications gradually increase after 2000 and patent numbers rapidly increase from 2005 to 2008. After 2008, patent application counts decrease year by year and stably development gradually.

Using prediction method of Pearl curve, this research calculates the counts of patent applications over the years to analyze the R&D trend and build the S-curve. The analytical result is shown in Fig. 4. From 1975 to 1993, both curves are slightly

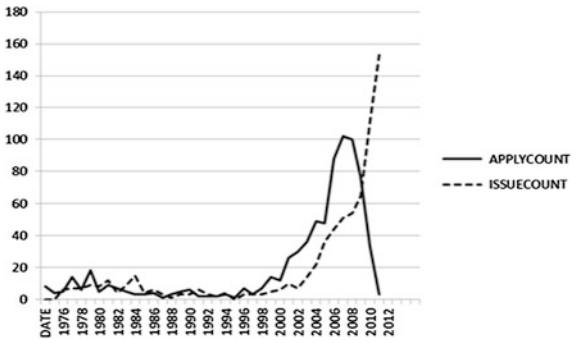


Fig. 3 Annual patents analysis

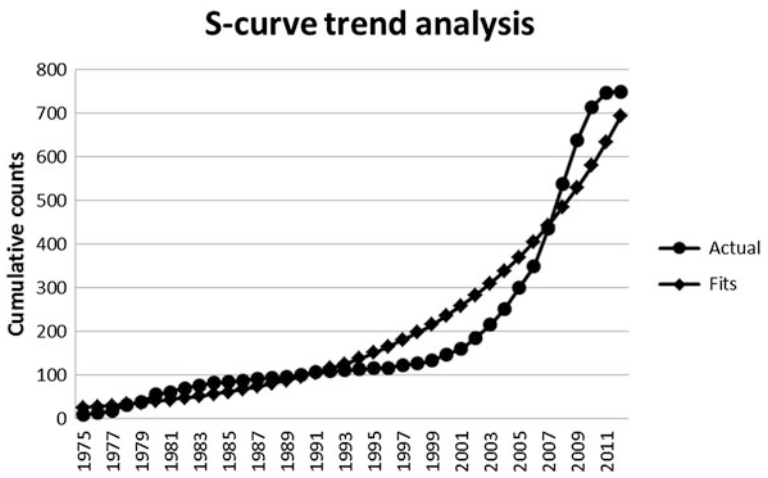


Fig. 4 S-curve trend analysis

different. After 1993, the historical data and predicted curve display gradual deviations. After 2005, actual curve grows rapidly and reaches stably development in 2012. On the other hand, the predicted S curve grows constantly from 1993 to 2012, and final cumulative counts are close to historical data in 2012.

Figure 5, shows most numbers of wind power related patents being owned by assignees from Europe, North America and East Asian. The leading countries in wind power technology are USA, Japan, Germany, Canada, and Denmark. There are 142 patents in USA, 111 patents in Japan and 76 patents in Germany. In Taiwan, there are eight related patents from Industrial Technology Research Institute (ITRI) and private companies, e.g., Hiwin Mikrosystem Corp.

The top three IPC (in third-order classification) are F03D, H02P and F03B, as shown in Table 2. F03D means wind power generator. H02P means motors, generators, electromechanical converter control, regulation, control transformers,

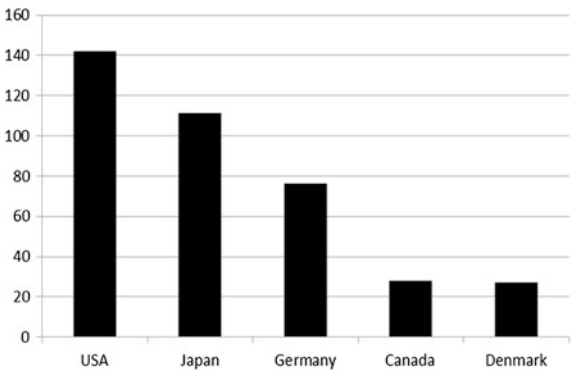


Fig. 5 Countries ranking analysis

Table 2 Top five third-order IPC patent counts

IPC	Patent counts
F03D	494
H02P	85
F03B	27
H02K	16
F01D	12

Table 3 Top three fifth-order IPC patent counts

IPC	Counts	Definitions
F03D9/00	271	Adaptations of wind motors for special use
H02P9/04	37	Control effected upon non-electric prime mover
H02P9/00	32	Arrangements for controlling electric generators

reactors or choke. F03B means hydraulic machine or hydraulic engine. Further, the top three IPC (in fifth-order classification) are F03D9/00, H02P9/04 and H02P9/00, as shown in Table 3.

General Electric Company (GE), the industry leader, owns patents belonging to IPCs of F03D11/00, F03D7/00, F03D7/04, F03D9/00, H02P9/00 and H02P9/04. GE focuses on F03D9/00 and H02P9/00. However, Mitsubishi Heavy Industries focuses on F03D9/00, F03D11/00 and H02P9/04. In technology distribution, GE leads development of wind power. Overall, the top two IPC are F03D9/00 and H02P9/04. We also draw IPC radar chart, as shown in Fig. 6.

Moreover, the patent forward citation rate analysis is an important basis to judge patent quality by the management analysis. This research conducts rating analysis for important patents, national patent quality and assignees. According to two

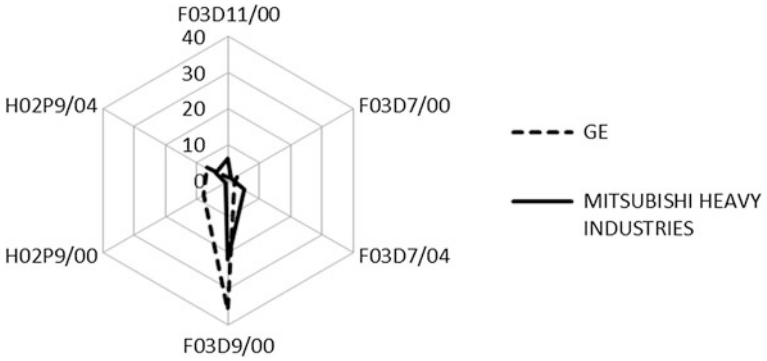


Fig. 6 Competitive company IPC radar chart

Table 4 Annual average forward citation rate

Patent number	Assignee	Country	Inventor	IPC	Year	Forward citation counts	Annual average forward citation rate
US7042110	Clipper Windpower Technology, Inc	Canada	Mikhail; Amir S.	H02P9/00	9	72	8
US6566764	Vestas wind systems A/S	Denmark	Rebsdorf; Anders V	F03D9/00	12	94	7.83
US5225712	U.S. Windpower, Inc.	Canada	Erdman; William L	F03D7/04	22	161	7.31
US7075189	Ocean wind energy systems	USA	Heronemus; Phyllis R	F03D9/00	10	73	7.3
US5083039	U.S. Windpower, Inc	Canada	Richardson; Robert D	F03D7/04	22	156	7.09

Table 5 Top 5 assignees patent quality rating

Assignee	Total forward citation counts	Patent number	Average forward citation rate	High forward citation counts	High forward citation rate (%)
General electric company (NY)	760	87	8.735	16	18.3
United Technologies Corporation (Hartford, CT)	500	10	50	10	100
Vestas wind systems A/S(DK)	299	23	13	5	21.7
Northern power systems, INC. (Waitsfield, VT)	143	8	17.875	5	62.5
U.S. wind power, INC (Livermore, CA)	471	4	117.75	4	100

citation indicators, such as total forward citation counts and annual average forward citation rate, we implement patent quality analysis.

This research collects the important patents in Table 4 that have greater than 7 times of annual average forward citation rate. The table contains the metadata of patent, such as assignees, inventors, IPC, forward citation counts and annual average forward citation rate. The life of patent No. US5225712 is 22 years, and it got a high forward citation that the domain engineer should analyze the claim of the important patent. Moreover, Table 5, the average of high forward citation rate is defined as 9.388 times. When forward citation counts of patent is larger than 9.388, it is called high forward citation patent. In high forward citation analysis, the enterprise, General Electric Company and United Technologies Corporation, are better than other assignees. General Electric Company owns the highest total forward citation and patent number, but high forward citation ratio is only 18.3 %. However, United Technologies Corporation owns 10 patents which are all of the high forward citation patents. Therefore, the company controls many key technologies in OWP domain and the related patents are worth for domain enterprises attention.

5 Conclusion

This research identifies some new findings through its ontology-based patent analysis. First, the main technology of offshore wind power (OWP) is represented based on technical domain ontology and all relevant patents are identified accordingly. Second, the key phrases of collected patents are extracted automatically and dynamically using text mining techniques and further verified by domain

specialists for an accurate OWP ontology schema. Third, using ontology-based patent analysis helps us understand the development of innovations within the OWP context, the technology life cycle, and the OWP technical leaders. The technology of wind power development has rapid growth between 2005 and 2009. The main countries, which own the most related patents, are USA (166), Japan (108), Germany (76), Canada (28) and Denmark (27). The important patent assignees include General Electric Company, Vestas Wind Systems A/S, Wobben, Aloys, Mitsubishi Heavy Industries, and Repower Systems AG. The IPCs of wind power are focused on three classifications, i.e., F03D, H02P and F03B. Particularly, F03D9/00 is the key technology category for developing adaptations of wind motors for special use and combinations of wind motors with apparatus driven. Based on the OWP ontology, this research collects the related patents and analyzes the patent assignees, forward citation rates, and the patent family information. The critical patents found are US7042110, US6566764, US5225712, US7075189 and US5083039. The patent US7075189 (offshore wind turbine with multiple wind rotors and floating system) is a high forward citation patent in the OWP domain. The patent indicates that wind energy conversion system is optimized for offshore application, each wind turbine includes a semi-submersible hull with ballast weight that is moveable to increase the system's stability, and the equipment associated with each rotor is located at the base of the tower to lower the metacentric height. Through the OWP patent analysis, Taiwan green policy makers and OWP energy supply chains can fast grasp the key technology leaders, critical patents, technical maturity cycles, and, most importantly, capture opportunities successfully in developing and implementing offshore wind power infrastructure.

Acknowledgments This research was partially supported by Taiwan National Science Council research grants (NSC-100-2221-E-007-034-MY3 and NSC-102-3113-P-027 -002), as well as the Ministry of Education grant for Advanced Manufacturing and Service Management Research at the National Tsing Hua University.

References

1. Andreasen T, Nilsson JF (2004) Grammatical specification of domain ontologies. Data Knowl Eng 48:221–230
2. Gao L, Porter AL, Wang J, Fang S, Zhang X, Ma T, Huang L (2012) Technology life cycle analysis method based on patent documents. Technol Forecast Soc Chang 80(3):398–407
3. Grandstrand O (1999) The economics and management of intellectual property: toward intellectual capitalism. Edward Elgar Publishing, London
4. Grilliches Z (1990) Patent statistics as economic indicators: a survey. J Econ Lit 28:1661–1707
5. Henry C The Growth Curve. Technology and Operations Management, California Polytechnic and State University. http://www.google.com.tw/url?sa=t&rct=j&q=pearl%2BCurve&source=web&cd=2&ved=0CC8QFjAB&url=http%3A%2F%2Fwww.csupomona.edu%2F~hco%2FMoT%2F03b_Growth_Curve.pptx&ei=YU7aUcXWNIWakAWD3YDQAw&usg=AFQjCNFneNAeyahrv9zGtP1kpp2WfEGVkQ

6. Leu HJ, Wu CC, Lin CY (2012) Technology exploration and forecasting of biofuels and biohydrogen energy from patent analysis. Int J Hydrogen Energy 37(20):15719–15725
7. Paci R, Sassu A, Usai S (1997) International patenting and national technological specialization. Technovation 17(1):25–38
8. Park H, Yoon J, Kim K (2013) Using function-based patent analysis to identify potential application areas of technology for technology transfer. Expert Syst Appl 40(13):5260–5265
9. Poli R (2002) Ontological Methodology. Int J Hum Comput Stud 56:639–664
10. Safer Environment (2008) Wind energy: renewable energy harnesses natural wind power—effective answer for emission problem towards cleaner, safer and greener environment. http://saferenvironment.wordpress.com/2008/11/03/wind-energy-renewable-energy-harnesses-natural-wind-power-%E2%80%93-effective-answer-for-emission-problem-towards-cleaner-safer-and-greener-environment/, Accessed: 05/28/2013
11. Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. Inf Process Manage 24(6):512–523
12. Scherer FM (1982) Inter-industry technology flows in the United States. Res Policy 11:227–245
13. Suzuki S (2011) Introduction to patent map analysis. Tokyo University of Agriculture and Technology. http://www.training-jpo.go.jp/en/uploads/text_vtr/pdf/Introduction%20to%20Patent%20Map%20Analysis2011.pdf
14. TIPO (2013) Taiwan patent terminology dictionary. http://paterm.tipo.gov.tw/IPOTechTerm, Accessed: 05/28/2013
15. Uschold M, King M (1995) Towards a methodology for building ontologies. In: IJCAI workshop on basic ontological issues in knowledge sharing, proceedings of the fourteenth international joint conference on artificial intelligence (IJCAI-95), Montreal, Quebec, Canada

Strategic Asset Management for Campus Facilities: Balanced Scorecard

Yui Yip Lau and Tsz Leung Yip

Abstract In the context of educational research, campus facility management is in the spotlight and plays a significant role in the twenty-first century. Campus facility management has led the school not only optimizing the running costs of buildings, but also increasing the efficiency and suitability of the management of space and other related assists management for people and processes. However, deficient uses of campus facilities continuously appear. Diseconomies of scale and lower service level are likely resulted. In order to find out an optimal solution for the campus facility planning for schools, this paper aims to illustrate capacity planning and control for the campus facility management and to further evaluate the associated logistics strategies. This paper presents a case study on a campus facility management, which is related to the proposed framework for performing efficiency and effectiveness at the different levels of planning including strategic, tactical and operational. Additionally, we also apply the balanced scorecard framework including rigorously different perspectives of financial, customer, internal operations, and continual learning and growth to evaluate the performance of campus facility management. Thus, the school could maintain the balance between the demand placed on a campus facility management and its ability to satisfy market demands in a cost-efficient way.

1 Introduction

In the past, Facilities management (FM) has been seen in the old-fashioned sense of cleaning, care-taking, repairs and maintenance. Thus, FM has been appeared in the poor relations within the architecture, real estate, construction professions and

Y.Y. Lau (✉)
Hong Kong Community College, The Hong Kong Polytechnic University, Hong Kong, China
e-mail: yylau@hkcc-polyu.edu.hk

T.L. Yip
Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong, China
e-mail: t.l.yip@polyu.edu.hk

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_145

engineering [8]. Since 1980s, FM has established itself as a key service sector [1, 14], for instance, real estate management, financial management, change management, human resources management, health and safety, contract management, building and engineering services maintenance, domestic services and utilities supplies. Nevertheless, FM covers a wide range of properties and user related functions [8]. FM should be positioned and included as a strategic support function of an organization (Ali and Mohamad [1, 5]. Several academic scholars (e.g., [2, 16, 18] emphasized the importance of the strategic role of FM to the core business. Cardoen et al. [3] has highlighted that 247 manuscripts have discussed about FM. Professional association and institutions such as The British Institute of Facilities Management and the International Facility Management Association have been established as separate disciplines in recognition of FM [8]. The International Facility Management Association currently defines FM as "a profession that encompasses multiple disciplines to ensure functionality of the built environment by integrating people, place, process and technology."

In the context of educational research, campus facility management is in the spotlight and plays a significant role in retaining and attracting students in a competitive demand—driven tertiary environment in the twenty-first century [9, 15]. Campus facility management is not only optimizing the running costs of buildings, but also increasing the efficiency and suitability of the management of space and other related assists management for people and processes [8]. Gow [6] noted that commencing students have based their decision to come to a particular school on the quality/appearance of its campus, its building and its facilities. However, deficient uses of campus facilities continuously appear in because:

The school operating capacity is fixed. However, the number of students is determined by external environment factors (i.e., uncontrollable elements e.g. demand uncertainty, fierce competition). The capacity cannot be adjusted easily to catch up with the change of external environment. Inaccurate forecasting leads to inefficient use of campus. In general, a school copes with excess capacity (i.e., 30 %) in the peak season and over capacity (i.e., 70 %) in the low season.

There is a long lead time (i.e., 2 years) to plan ahead of the programme operations. Initially, the school needs to conduct marketing research to evaluate the demand of programme for half year. Then, the programme needs to obtain both home country and Institute Senate approval for one year. Finally, the programme will be reviewed under Non-Local Programme Registration (Cap 493) at Hong Kong for half year. Long lead time deviate the original allocated facilities of each programme. Hence, the school faces the risk of under-estimation or over-estimation of capacity.

Diseconomies of scale and lower service level are likely resulted. In order to find out an optimal solution for the campus facility planning for schools, this paper aims to illustrate capacity planning and control for the campus facility management and to further evaluate the associated logistics strategies. This paper presents a case study on a campus facility management, which is related to the proposed framework for performing efficiency and effectiveness at the different levels of planning including strategic, tactical and operational. Furthermore, we illustrate the balanced

scorecard framework [12] including rigorously different perspectives of financial, customer, internal operations, and continual learning and growth to evaluate the performance of campus facility management. Thus, the school could maintain the balance between the demand placed on a campus facility management and its ability to satisfy market demands in a cost-efficient way.

2 Case Investigation

In this study, we illustrate a case study to discuss how the school adopts logistics strategy in campus facility management. There are 7 classrooms (i.e., classroom 1 and 2 are served for 39 students; classroom 3, 4 and 5 are served for 3 students; classroom 6 is served for 25 students; classroom 7 is served for 13 students) with 125 students' capacity. The floor plan of school is illustrated in Fig. 1. The school needs to forecast the number of students in 8 high educational programmes (i.e., from higher certificate to undergraduate degree to master degree), 2 main kinds of tutorial classes (i.e., IGCSE, GCE) with different modes of study (full time, part time).

2.1 Strategic Level

The school considers location where it has highly accessiblity and synergy effect. Highly accessiblity can attract students to study in a comfortable environment. Moreover, the school is able to utilize their resources and maximize their profit under synergy effect. For instance, the school considers their clusters in the same district areas so as to reduce operating cost and increase efficiency.

For the school size, we need to maximum operating capacity (efficiency goal) to cope with the changing demand [10]. According to Johansen [7], capacity is the

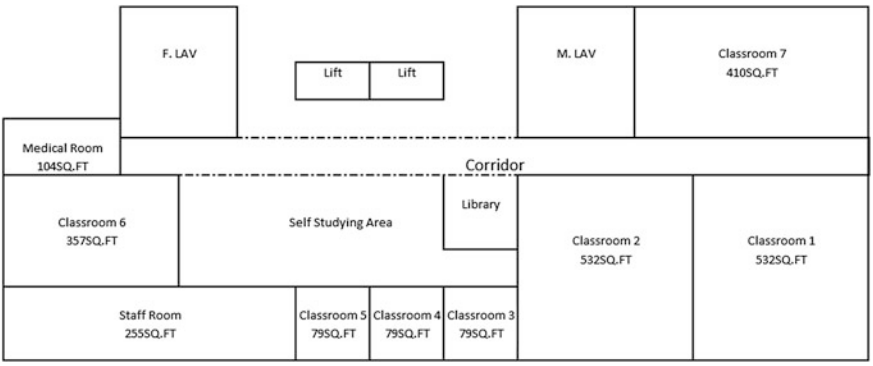


Fig. 1 School floor plan

maximum amount that can be produced per unit of time with the existing plant and equipment, provided that the availability of variable factors of production is not restricted. In the current situation, the school is faced capacity allocation problem which is the problem of finding an optimal allocation of capacity to meet each demand [4]. There are technical constraints for the capacity allocation problem due to the numbers of students is determined by external environment while the school is in fixed operating capacity. Thus, a school is faced with over capacity in the low season, as well as excess capacity in the peak season.

Over-capacity: The school leases out the classrooms for other institutions to organize the activities including seminars, workshops, exhibitions, information sessions and job interviews. Additionally, a school collaborates with associations to deliver short courses. Hence, the classroom utilization rate would be significantly increased.

Excess capacity: In economic and engineering terms, excess capacity means that a school is failure to reach potential output [17]. Melnick et al. [11] found that the school encounter financial distress and may have to modify the operations. One way to modify operations has been to rent the centre which is approved by Education Bureau. Alternatively, the school established the branches under provisional registration (Cap 279). This could provide reliable learning space without determined by other parties. However, the procedure of school registration is complicated as well as prepared for long lead time in registration period for 4–6 months.

2.2 Tactical Level

The school designs the tactical level strategy which usually has 1–2 year time frames. Based on the demand for education, the school adjusts the workforce size and campus facilities accordingly. For instance, there is significant increasing demand for higher education in 2012. In 2012, there were double cohort combining the last Hong Kong Advanced Level Examination candidates and the first Hong Kong Diploma of Secondary Education (HKDSE) Examination candidates. The school targeted a large pool of potential students whose are looking for our Higher Diploma programmes. In order to producing services as efficiently as possible within the strategic plan, the following measures have been taken:

Recruit a team of qualified teaching staff in Business Administration area for teaching Institution C (three Higher National Diploma programmes) and Institution D (one Higher National Certificate programme and one Higher National Diploma programme).

Recruit a team of Quality Assurance (QA) and External Examiners/Assessors to maintain the quality of teaching staff, programmes, facilities and learning materials.

Provide additional administrative support specially to tailor made for students to take Higher Diploma programmes like adding administrative staff, establishing library, extending opening hours for students to submit assignments and do revision, reserving separate classrooms for each programme.

2.3 Operational Level

For operational level strategy, the school designs the timetable for each programme to deliver support for routine needs. Programme characteristics are described in Table 1.

From the operational perspective, the school implements the timetable have encountered the problem of assigning a number of programmes to a given number of timeslots and rooms while satisfying a set of constraints [13]. At this stage, we normally plan for one semester. In design and plan the timetable, we need to consider the critical factors pertaining to mode of study, number of students, number of subjects in one semester, studying schedule and required studying hours per subject so as to ease of implementation and their flexibility. After that, we allocate the qualified teaching staff and classroom for different programmes and courses in different time slots. This could help the school to understand the utilization rate of different classrooms and the demand for different programmes and courses. Furthermore, it is easy for the school to control their daily operations. As presented in Table 2, the school allocates different classrooms which are based on programme characteristics.

3 BSC Application in Campus Facility Management

In the context of campus facility management, BSC has evolved to become an effective strategy execution framework to improve internal and external communications, monitor the performance against strategic goals and align school activities to the vision and strategy of the school.

Table 1 Programme details

Programme	Institution	Level of study	Mode of study	Study period (months)
1	A	Master	Part time	24
2	B	Master	Part time	20
3	B	Master	Part time	24
4	C	Higher National Diploma	Full time	18
5	C	Higher National Diploma	Full time	18
6	C	Higher National Diploma	Full time	18
7	D	Higher National Certificate	Full time	12
8	D	Higher National Diploma	Full time	12

Table 2 Summary of programmes in classroom allocation

Programme	Reasons of allocation
1	The school charges students' tuition fee is the most expensive. Both expectations of students and institutions in the school facilities are high. The school assigns this master programme students use in classroom 1. The reasons are:
	The largest size of this classroom makes students have a comfortable learning space
	The majority of students are positioned at Executive and Senior Management levels same as teaching and learning facilities are provided
	This classroom could be served as a spare area for students to have a break with snack, coffee and wine
	Institution A assigns classroom 1 to operate this master programme. They request the school to demonstrate unique Scottish characteristics in this classroom pertaining to Scottish pictures (i.e., landscape, culture, traditional buildings) and Scottish souvenirs
	Classroom 1 is located near classroom 7 (i.e., computer room). It is convenience for students to do assignments and access electronic networks
2	The school charges students' tuition fee is expensive. The expectations of students in the school facilities are high. The largest size of this classroom makes students have a comfortable learning space
	This classroom could be reserved as a spare area for students to have a break with snack and coffee
	Students need to take self-studying for 100 hours in this programme. Thus, accordingly, the school assigns this master programme students use in classroom 2 as this classroom is located near self-studying area and library. It is convenience for students to do assignments, conduct self-learning and participate in group discussion
3	The school charges students' tuition fee is expensive. The expectations of students in the school facilities are high. The largest size of this classroom makes students have a comfortable learning space
	This classroom could be reserved as a spare area for students to have a break with snack and coffee
	Students need to take self-studying for 120 hours in this programme. The school has established a small scale of library in this classroom. Hence, students could conduct self-learning and assignments
	This master programme has included one subject in Practical Nursing Studies. Medical room is located near at classroom 6. It provides synergistic effect in programme operations
4	This programme aims to target for 25 students. The objective is to keep 100 % capacity utilization rate
5	This programme aims to target for 39 students. The objective is to maintain 100 % capacity utilization rate
6	This programme aims to obtain 39 students. The objective is to reach at 100 % capacity utilization rate

(continued)

Table 2 (continued)

Programme	Reasons of allocation
7	This programme aims to obtain 25 students so as to keep 100 % capacity utilization rate
	Due to other programmes have occupied all weekdays session (i.e., before 1430), Saturday and Sunday, this programme is only allocated in all weekdays sessions (i.e., after 1430)
8	This programme aims to have 39 students in order to maintain 100 % capacity utilization rate
	This programme is only allocated in all weekdays sessions (i.e., after 1430) because of other programmes have used all weekdays session (i.e., before 1430), Saturday and Sunday

In summary, we apply the BSC framework consists of rigorously four main perspectives pertaining to financial, customer, internal operations, and continual learning and growth. Sixteen evaluation sub-criteria are embedded in the four main perspectives. The conceptual linkage between the campus facility management and the BSC is demonstrated in Fig. 2.

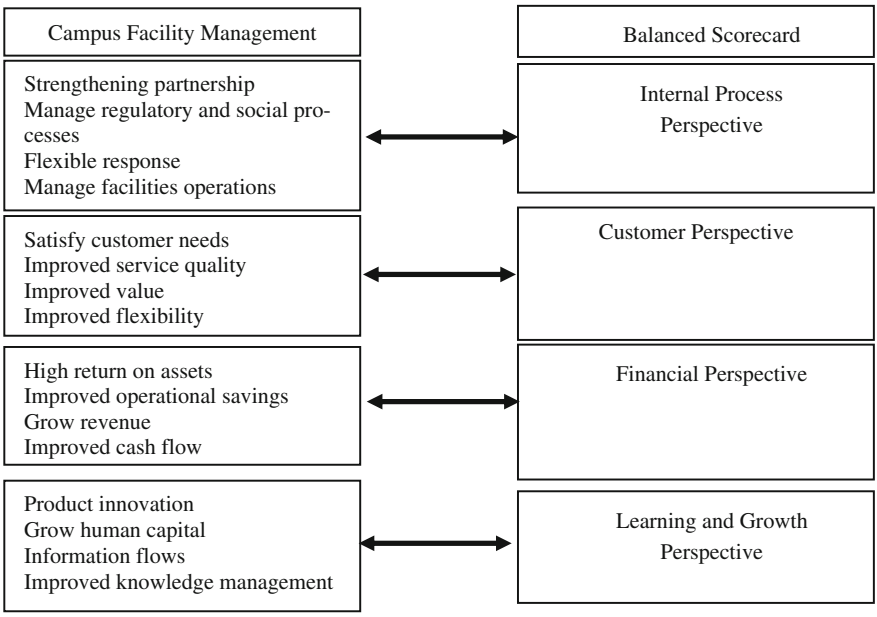


Fig. 2 Linking campus facility management to balanced scorecard

4 Discussion

4.1 Learning and Growth Perspective

For the growth of human capital, we need to increase staff retention and job satisfaction index (i.e., staff turnover, job security, absenteeism, complaints and work environment). For instance, the popular teaching staff and qualified supporting staff is a human capital of campus facility management. They could attract the students to take the programmes. Both of them have a high job satisfaction rate leads to a high chance of retention. Moreover, an installation of information systems in campus in which facilitates the flow of information in various departments and maintain smooth operations. In addition, the school has established the training department to design the in-house training programmes regarding to campus facility management. Additionally, the number of in-house training and learning hours has significantly increased from 5 to 10 hours per week. Knowledge management would be improved. Furthermore, we need to develop the trendy and customized education programmes in order to encourage students to study at our campus. This is a form of product innovation. Hence, it could bring a substantial growth in campus facility development.

4.2 Internal Process Perspective

In order to facilitate the internal process in campus facility management, we could summarize into the following four major dimensions. Firstly, we need to find partners whose are active, reliable and high partner confidence score. This could strengthen partnership relationships in the long term. Secondly, the campus requires complying with the laws and regulations with Education Bureau and Hong Kong Council for Accreditation of Academic and Vocational Qualifications (HKCAAVQ) so as to improve the alignment between school operations and external parties. Thirdly, the campus facilities need to maintain flexible response according to the demand for the programmes and the request of students. For instance, the traditional classroom has converted into computer room due to the emergence of Information Technology, Creation Media, Interior Design and Visual Communications programmes. Finally, the campus needs to manage facilities operations including the utilization rate, allocation of students in different classrooms, inflow and outflow of students in a campus, and integrity of campus facilities.

4.3 Customer Perspective

From the perspective of campus facility management, the school needs to deliver the service quality not only satisfy student needs, but also exceed student's expectations. For instance, responding to the student's enquires in a real time, extending opening hours and convert small classrooms for students to do revision during an examination period, leasing the computer notebooks to students and purchasing additional reference books. All the above student-centric measures could improve service quality, value and flexibility, as well as reduce the number of complaints from students.

4.4 Financial Perspective

The dual objectives of facility management are cost minimization and profit maximization. For cost minimization, the school needs to reduce operating expenses (i.e., teaching staff could take teaching duties with additional administrative tasks; save energy; set up online administrative system to reduce manpower support). Also, the school could increase operational efficiency in order to reduce operational cost and improve cash flow. For profit maximization, the school could develop a wide variety of educational programmes (i.e., from Higher National Certificate to Master Degree) aims to increase the number of student enrolment. On the other hand, the school could optimize resource utilization so as to obtain high return on assets. The school would implement the following measures:

The empty timeslot could be replaced by GCE and IGCSE tutorial classes. GCE and IGCSE tutorial classes are easy to commence due to (1) flexibility in changing the timeslot in short notice (i.e., 2 dates in advance); (2) a free of minimum number of student requirements; (3) students are allowed to choose any subject in tutorial classes rather than taking the whole programme. Students could reduce their financial burden and risk. Sometimes, the small class size of GCE and IGCSE tutorial classes (i.e., 8 students or below) use the large classrooms due to all small classrooms are fully booked, as well as large classrooms have not occupied by any programmes or activities. The school faced a problem of low utilization of facilities.

The school leases out the classrooms for other institutions to organize the variety of activities like seminars, workshops, exhibitions, information sessions and job interviews. It not only increases the classroom utilization rate, but also promotes the image to others.

The school collaborates with associations and corporates to deliver short courses. Hence, the classroom utilization rate would be significantly increased.

There are three small classrooms (i.e., classroom 3, 4 and 5). It could be converted from teaching rooms into meeting/consultation rooms within a short time. The school could develop new business area (i.e., overseas student recruitment). The overseas institutions could rent these kinds of meeting/consultation rooms in a

specific time so as to recruit students whose would like to study ranging from Boarding School, Foundation Degree, Undergraduate, Postgraduate and Doctorate degree overseas.

The school has installed computers in classroom 7. It could provide additional administrative support for students to do their assignments and revision. Classroom 7 could be easily changed into computer room for multi-purposes. Thus, the school could develop new courses like Creation Media, Interior Design and Visual Communications.

5 Summary and Conclusions

In our analysis, it is worth to discuss what managerial implications are affected to the campus facility management. Our key focus on (1) Improve efficiency and effectiveness at the different levels of planning pertaining to strategic, tactical and operational; (2) Exploit key capabilities and core competence effectively; and (3) Improve key weakness diligently. The sixteen evaluation sub-criteria could facilitate the school to transform into potential key capabilities in campus facility management.

References

1. Adewunmi Y, Ajayi C, Ogunba O (2009) Facilities management: factors influencing the role of Nigerian estate surveyors. J Facil Manag 7(3):246–258
2. Atkin B, Brooks A (2000) Total facilities management. Blackwell Science, London
3. Cardoen B, Domeulemeester E, Belien J (2010) Operating room planning and scheduling: a literature review. Eur J Oper Res 201:921–932
4. Cao C, Gao Z, Li K (2012) Capacity allocation problem with random demands for the rail container carrier. Eur J Oper Res 217:214–221
5. Chotipanich S, Lertariyanun V (2011) A study of facility management strategy: the case of commercial banks in Thailand. J Facil Manag 9(4):282–299
6. Gow G (1999) Shifts in the campus planning and development paradigm. In: ATEM Conference, Wellington, New Zealand
7. Johansen L (1968) Production functions and the concept of capacity. In: Fo'rsund FR (ed) Collected works of Leif Johansen. North-Holland Press, Amsterdam, p 987
8. Kamaruzzaman SN, Zawawi EMA (2010) Development of facilities management in Malaysia. J Facil Manag 8(1):75–81
9. McLaughlin P, Faulkner J (2012) Flexible spaces: what students expect from university facilities. J Facil Manag 10(2):140–149
10. Medernach E, Sanlaville E (2012) Fair resource allocation for different scenarios of demands. Eur J Oper Res 218:339–350
11. Melnick G, Keeler E, Zwanziger J (1999) Market power and hospital pricing: are nonprofits different? Health Aff 18(3):167–173
12. Michalska J (2005) The usage of the balanced scorecard for the estimation of the enterprise's effectiveness. J Mater Process Technol 162–163:751–758

13. Sabar NR, Ayob M, Kendall G, Qu R (2012) A honey-bee mating optimization algorithm for educational timetabling problems. Eur J Oper Res 216:533–543
14. Scupola A (2012) Managerial perception of service innovation in facility management organizations. J Facil Manag 10(3):198–211
15. Temple P (2008) Learning spaces in higher education: an under-research topic. London Rev Educ 6(3):229–241
16. Then DSS (2003) Strategic management. In: Best R, Langston C, de Valence G (eds) Workplace strategies and facilities management. Butterworth-Heinemann, New York, pp 69–80
17. Valdmanis V, Bernet P, Moises J (2010) Hospital capacity, capability, and emergency preparedness. Eur J Oper Res 207:1628–1634
18. Varcoe B (2000) Implications for facility management of the changing business climate. Facilities 18:383–391

Bayesian Approach to Determine the Test Plan of Reliability Qualification Test

Kun Yuan and Xiao-Gang Li

Abstract In view of the shortage of large sample size in Reliability Qualification Test, a Bayesian model is introduced in this paper. Taking advantage of the historical information and expert experience as the prior information, the reliability of the product at current stage is estimated by making use of the new Dirichlet distribution and filed test information in reliability growth testing. Because of the parameters of the new Dirichlet distribution having no physical meaning, to obtain the parameter estimators of prior distribution, we adopt the ML-II method to transfer the failure information at different stages. Considering the balance of consumer's risk and producer's risk, the minimum testing plan is scheduled. This method could be regarded as a supplementary program in reliability testing for qualification. An example illustrates the contrast testing test sample size between the proposed method and the standard scheme which validate the proposed approach.

Keywords Reliability qualification test · Reliability growth · Bayesian analysis · New Dirichlet distribution · Test plan determination

1 Introduction

The aim of reliability qualification test (RQT) is to examine whether the designed product has met the reliability requirement or not. RQT helps achieve overall system reliability, and it can be used as a foundation to finalize the design. But according to the standard RQT requires a significant investment and a long-term commitment to run and maintain them. This drawback limits the extensive use of

K. Yuan (✉) · X.-G. Li
Beijing University of Aeronautics and Astronautics, Beijing, China
e-mail: yuankun@dse.buaa.edu.cn

X.-G. Li
e-mail: lxg@buaa.edu.cn

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_146

RQT and even is impossible to those complex systems. However, during the design phase of a product, especially to the complex system, there is always a process consisted of design, test, modification, and retest. In every stage massive amounts of testing data can be collected. Thus we can utilize the prior information inferred from the field data and expert information to reduce the test time and the number of products in RQT.

During the design phase, the reliability growth test (RGT) is adopted to improve the reliability of the product through testing and modification. Taking full advantage of prior information in RGT could substantially reduce the time and money in RQT. Much work has been done in the area of statistical analysis for RGT [1]. There are many models which are proposed such as Smith model [2], Barlow-Scheuer model [3], and Logistic model and so on. These models are proposed based on non-informative prior distribution, so that the reliability in current stage could be inferred, but the final reliability of the product can't be obtained. Mazzuchi and Soyer [4, 5] firstly introduce a family of prior distribution, the ordered Dirichlet distribution, for binomial trials to evaluate the reliability in RGT. The advantage of this distribution is the combination of the posterior expression for joint and marginal distributions and the historical information as well as expert experience. Reliability estimation and prediction could be inferred by making use of the model. Erkanli [6] extends this model to sequential test. Liu fei apply the Bayesian method to RGT for exponential life with the prior distribution of Dirichlet. However, the proposed models are based on the ordered Dirichlet prior distribution and Li [7] points out the shortage of this Bayesian method. He propose a new model called new Dirichlet distribution, which similar to the ordered Dirichlet prior distribution, but in every stage of the test there are two parameters to describe the reliability estimation and degree of certainty. This model is obviously better to describe the reliability estimation in RGT than the ordered Dirichlet distribution. Wu [8] extends the new Dirichlet prior distribution to exponential life product. However, another problem is raised that the physical meaning of the parameters in every stage are indefinite so that they can't be assigned with expert information.

In view of the long test time in RQT, a Bayesian method is proposed by utilizing the inferred reliability in RGT and expert information. Zhang [9] proposes a Bayesian plan for exponential case which synthesizes historical data in RGT, and then he [10] takes advantage of engineering experience to obtain the parameter estimators of Beta prior distribution to determine the test plan of RQT. Ming [11] introduces the mixed Beta distribution as the prior distribution and inheritance factor to infer the Bayesian plan. But these proposed method doesn't take full advantage of the failure information at different stages since they ignore the reliability relationship between every two stages. On the other hand, the obtained parameter estimators of the prior distribution were influenced by subjectivity which may lead to inaccuracy. So in this paper we introduce the new Dirichlet distribution as prior distribution which describes the reliability relationship clearly in RGT. The prior distribution of the reliability is calculated with ML-II method by making use of prior information and field test information at every stage. Then we considered the modified inferred posterior distribution in RGT as prior distribution in RQT,

and the test plan is obtained through Bayes inference. Finally, a numerical example is presented to compare the proposed method to the test plan in standard, to demonstrate the validity of this method.

2 Reliability Growth Model

Assumption (reference [2])

(a) Trials are statically independent, and during stage i the probability of a success is R_i,$i = 1, 2, \cdots, m$.
(b) During stage i, there are n_i trials $(n_i \geq 1)$ resulting in s_i successes$(0 \leq s_i \leq n_i)$.
(c) Due to modification, $R_1 \leq R_2 \leq \cdots \leq R_m$.

Then the likelihood function in RGT can be presented as follows

$$L(R_i; n_i, s_i) = \binom{n_i}{s_i} R_i^{s_i} (1 - R_i)^{n_i - s_i} \tag{1}$$

R_i is unknown and n_i, s_i is known, $i = 1, 2, \cdots, m$.

Considering the weak point of the ordered Dirichlet prior distribution, Li [7] proposed a new family of prior distributions. This prior distribution adopted the conditional form and the linearly transformed beta distribution, which could overcome the shortage of the ordered Dirichlet prior distribution and conveyed the engineer's prior knowledge clearer.

During stage i, a Beta distribution in $(R_{i-1}, 1)$ is constructed to estimate the reliability at current stage with conditional prior form, that is

$$\begin{aligned} g_i(R_i|R_{i-1}) &= g_i(R_i|R_{i-1}; a_i, b_i) \\ &= [B(a_i, b_i)]^{-1} (R_i - R_{i-1})^{a_i - 1} (1 - R_i)^{b_i - 1} (1 - R_{i-1})^{1 - a_i - b_i} \end{aligned} \tag{2}$$

where $a_i > 0$, $b_i > 0$, $R_0 = 0$ and $R_{m+2} = 1$. $g_i(R_i|R_{i-1}, a_i, b_i)$ is Beta function defined in $(R_{i-1}, 1)$ which is similar to the standardized Beta function, and $B(a_i, b_i)$ can be wrote as the following form with *Gamma* function

$$B(a_i, b_i) = \frac{\Gamma(a_i)\Gamma(b_i)}{\Gamma(a_i + b_i)} \tag{3}$$

Note that if we set $R_0 = 0$, then g_1 can be expressed as following form

$$g_1(R_1) = [B(a_1, b_1)]^{-1} (R_1 - R_0)^{a_1 - 1} (1 - R_1)^{b_1 - 1} (1 - R_0)^{1 - a_1 - b_1} \tag{4}$$

Which is the same form as g_i in (2).

During every stage in RGT, two different parameters are used to describe the reliability growth, and the expert information can also be more suitably delivered. With prior information inferred in the former stage, the prior information at current stage after modification is estimated. The procedure to obtain the prior distribution parameter is clearly depicted in the following part.

3 Determination of New Dirichlet Distribution Parameters

For given $R_0, R_1, \cdots, R_{k-1}$, the conditional mean and variance of R_k are, respectively,

$$\mu^* = E(R_k) = \frac{a_k + b_k R_{k-1}}{a_k + b_k} \tag{5}$$

$$v_k^* = Var(R_k) = \frac{a_k b_k (1 - R_{k-1})^2}{(a_k + b_k)^2 (a_k + b_k + 1)} \tag{6}$$

From the above formula it's evident that the physical meaning of the prior distribution parameters are not distinct. So only the determination of new Dirichlet distribution parameters based on prior information is solve, can we apply this method to the reliability estimation and prediction.

During every stage of the RGT, the prior distribution of the current stage product cannot be replaced by the posterior distribution of the previous stage because of modification introduced. So when the failure information at the former stage is transformed, we can apply it to gain the prior distribution at current stage. Numerous methods for example, conversion factor method, have been proposed. Hence here we adopt the ML-II method to transfer the failure information between stages. Details have been introduced by Mao [12].

Assume that at first stage, since no more information is informed, we utilize the uniform distribution as the prior distribution, which is $\pi_1(R_1) = 1,\ 0 \le R_1 \le 1$. After the first stage test, we obtain the test data(n_1, s_1), which s_1 denotes the number of successful products. So the posterior distribution of R_1

$$\pi_1(R_1|(s_1, n_1)) \sim Be(s_1 + 1, n_1 - s_1 + 1) \tag{7}$$

But the posterior distribution cannot directly employed as the prior distribution. Modification must be adopted for the parameter in distribution to transfer the failure information. We assume that

$$\pi_2(R_2) \sim Dir(s_1 + \alpha_1, n_1 - (s_1 + \alpha_1)) \tag{8}$$

Which α_1 denotes the correction.

Based on the Bayesian theory, we consider the test data (n_2, s_2) as samples of the marginal density distribution $m(n_2, s_2)$.

$$m(n_2, s_2) = \int_{R_1}^{1} \pi_2(R_2) R_2^{s_2} (1 - R_2)^{n_2 - s_2} dR_2 \tag{9}$$

where

$$\pi_2(R_2) = \frac{1}{B(s_1 + \alpha_1, n_1 - (s_1 + \alpha_1))} \left(1 - \tilde{R}_1\right)^{1-n_1} \left(R_2 - \tilde{R}_1\right)^{s_1 + \alpha_1 - 1} (1 - R_2)^{n_1 - (s_1 + \alpha_1) - 1} \tag{10}$$

Since the new Dirichlet distribution is on the condition of the former reliability, it implies the current reliability values in the range of R_{k-1} and 1, which means we have great confidence to ensure the estimates of reliability is correct. However, great risk has involved when we apply this method, so we set $\tilde{R}_1 = 0.1R_1$ to reduce this hazard. Then we get

$$\begin{aligned} m(n_2, s_2) = & \frac{1}{B(s_1 + \alpha_1, n_1 - (s_1 + \alpha_1))} \\ & \times \int_{R_1}^{1} \left(1 - \tilde{R}_1\right)^{1-n_1-n_2} \left(R_2 - \tilde{R}_1\right)^{s_1+s_2+\alpha_1-1} (1 - R_2)^{n_1+n_2-(s_1+\alpha_1+s_2)-1} dR_2 \end{aligned} \tag{11}$$

Method of approximation is applied to obtain the $\hat{\alpha}_1$ which maximizes the function $m(n_2, s_2)$. Then the prior distribution $\pi_2(R_2)$ can be expressed as the following form

$$\pi_2(R_2) = \frac{1}{B(s_1 + \hat{\alpha}_1, n_1 - (s_1 + \hat{\alpha}_1))} \left(1 - \tilde{R}_1\right)^{1-n_1} \left(R_2 - \tilde{R}_1\right)^{s_1 + \hat{\alpha}_1 - 1} (1 - R_2)^{n_1 - (s_1 + \hat{\alpha}_1) - 1} \tag{12}$$

With this method the prior distribution of every stage can be obtained. Go round the procedure introduced in this section we obtain the final reliability estimation at the end of the RGT step by step. The following part illustrates how the test plan is computed.

4 Identification of Test Plan in RQT

Assume that the specified acceptable quality level is p_0, and the limiting quality level is $p_1(p_1 < p_0)$. We set the test plan that there are n trials and the maximum failure is c. Then the prior distribution $\pi_k\left(R_k\middle|\tilde{R}_{k-1}; a_k, b_k\right)$ in RQT is inferred on the basis of modified posterior distribution in RGT and likelihood function as follow form

$$f\left(R_k\middle|\tilde{R}_{k-1}; a_k, b_k\right) = Dir\left(a_k + n - c, b_k + c\middle|\tilde{R}_{k-1}\right) \tag{13}$$

It's obvious that the posterior distribution $f(R|n, c)$ follows the Beta distribution. As inferred by Zhang [10], we gain the consumer's maximum posterior risk

$$\max_{s \geq n-c} P(p < p_1|s) = P(p < p_1|s = n - c) \tag{14}$$

If given the consumer's risk β, then we get

$$P(p < p_1|s = n - c) \leq \beta \tag{15}$$

On the other hand, the produce's maximum posterior risk

$$\max_{s \leq n-c-1} P(p \geq p_0|s) = P(p \geq p_0|s = n - c - 1) \tag{16}$$

If given the produce's risk α, then we get

$$P(p \geq p_0|s = n - c - 1) \leq \alpha \tag{17}$$

Only require that

$$P(p \geq p_0|s = n - c) \leq \alpha \tag{18}$$

This is Equivalent to the following form

$$P(p \leq p_0|s = n - c) \geq 1 - \alpha \tag{19}$$

Based on the above analysis, the key work to determine the test plan in RQT is to solve the following formulas when given the acceptable quality level p_0, the limiting quality level p_1, the consumer's risk β, and the produce's risk α to obtain n and c

$$\begin{cases} L(p_0) = P(p \leq p_0|s = n - c) \geq 1 - \alpha \\ L(p_1) = P(p \leq p_1|s = n - c) \leq \beta \end{cases} \tag{20}$$

where n and c should be integer

$$L(R) = P(p \le R_0 | s = n - c) = \int_{R_m}^{R_0} h(R|n, c)dR \tag{21}$$

When the product tested in RQT, what we concern is posterior risk of both consumer and producer. In this paper we utilize the Bayesian theory and new Dirichlet prior distribution to obtain a reasonable test plan in RQT, which less the number of test products and time compared to the test plan in standard.

5 Numerical Example

Consider an example that the specified acceptable quality level is 0.97, and the limiting quality level is 0.91 with the discrimination ratio is set to 3. Assume that the consumer's risk and the produce's risk are no more than 0.2. According to standard GJB899 we select (47.2) as the test plan, which means there are totally 47 products tested in RQT and no more than 2 product fail. The posterior risk we calculate are 0.183 and 0.17.

Assume that the product has been tested through four stages. We utilize the method mentioned in Sect. 2 to obtain the conditional prior Dirichlet distribution just presented in Table 1. In order to compare to test standard plan, we set $c = 2$ and solve the Eq. (17) to gain the number of test product n in Table 1.

From the result compared to the standard test plan, the advantage of the proposed method highlights and it's obvious to reduce the number of test product and gain evident economic benefit. Here for the consideration of risk, we set the interval $\left(0.1\tilde{R}_{k-1}, 1\right)$ instead of $\left(\tilde{R}_{k-1}, 1\right)$. If we are sure of the validity of estimation of $\tilde{R}_{k-1}$, the less testing number are required compared to this computation result. What we can also infer from the computation result is that the consumer's risk β is much bigger than the contrast produce's risk α, since the produced product have been approved by the consumer. A balance is needed to spare the risk to both sides.

Table 1 computation of new Dirichlet distribution

Stage k	(n_k, s_k)	α_k	a_k	b_k	Test plan	Actual risk		Saving ratio (%)
						α	β	
1	(4,3)	0	1	1	(47,2)	0.183	0.17	0
2	(5,4)	0.23	3.23	0.77	(38,2)	0.194	0.183	19.14
3	(6,5)	0.654	7.884	1.116	(37,2)	0.161	0.188	21.28
4	(7,6)	0.606	13.49	1.51	(35,2)	0.132	0.196	25.53

6 Conclusion

This article proposes a Bayes method, which is constructed on the basis of new Dirichlet prior distribution to determine the test plan of RQT. Since the new Dirichlet distribution is constructed in the interval of reliability in the former stage and 1, the advantage is evident to lessen the number of tested product. This method takes full advantage of reliability information and field test data. Some revision can be made to improve the accuracy of the model, such as extending the interval of reliability at current stage. The insurance of estimated reliability is also considered with the probability P_0 and P_1. We establish equations to obtain the test plan concerning both sides risk. The numerical example illustrate the practicality and economic benefit.

References

1. Teneja VS, Safie FM (1992) An overview of reliability growth models and their potential use for NASA application. NASA Technical Paper 3309
2. Smith AFM (1977) A Bayesian note on reliability growth during a development testing program. IEEE Trans Reliab 26:346–347
3. Barlow RE, Scheuer EM (1966) Reliability growth during a development testing program. Technometrics 8:53–60
4. Mazzuchi TA, Soyer R (1992) Reliability assessment and prediction during product development. In: Proceedings annual reliability and maintainability symposium, pp 468–474
5. Mazzuchi TA, Soyer R (1993) A Bayes methodology for assessing product reliability during development testing. IEEE Trans Reliab 42:503–510
6. Erkanli A, Mazzuchi TA, Soyer R (1998) Bayesian computations for a class of reliability growth models. Technometrics 40:14–23
7. Guo-ying LI, Qi-guang WU, Yong-Hui ZHAO (2002) On Bayesian analysis of binomial reliability growth. Japan Statist. Soc. 1:1–14
8. Qi-guang WU, Guo-ying LI, Yong-hui ZHAO (2003) A new family of a prior distributions for exponential reliability growth models. Acta Math Sci 3:474–484
9. Zhi-hua ZHANG, Li-ping JIANG (2000) A Bayesian plan of testing for production acceptance in exponential case. Chin J Appl Prob Stat 16(1):66–70
10. Li-ping JIANG, Zhi-hua ZHANG (2000) A Bayesian method of reliability qualification test in binomial case. J Eng Math 17(4):25–29
11. Zhimao Ming, Junyong Tao (2008) A Bayes plan of reliability qualification test based on the mixed Beta distribution for success/failure product. Acta Armamentarii 29(2):204–207
12. Shisong Mao (1999) Bayesian statistics. China Statistics Press, Beijing

Consolidating People, Process and Technology to Bridge the Great Wall of Operational and Information Technologies

Anastasia Govan Kuusk and Jing Gao

Abstract An organisations competitive advantage depends upon timely and consolidated provision of information to enable strategic decision making. Timely and consolidated provision of information requires integration of people, processes and technology across an organisation. In organisations that manage infrastructure assets such as power, water, sewerage, telecommunications or transport, timely and consolidated provision of information is impeded by a divide of operational and business people, processes and technology. Business areas such as finance may be supported by an Information Technology (IT) branch with well developed governance processes and dedicated information technology people such as network analysts, helpdesk and systems administrators. Operational areas such as power generation do not share such people, processes and technology. This chapter provides asset infrastructure organisations with recent research results indicating phases for consolidating Operational Technology (OT) and IT, identification of OT and IT divide and how to overcome the divide.

1 Introduction

Engineering Asset Management (EAM) manages risks when maintaining and replacing critical built asset infrastructure such as water, transport, sewerage and power services. Frameworks to manage such assets are important to the Australian economy as the estimated value of built assets in Australia in 2010 was $600 billion [1]. The study of EAM has evolved in the last 30 years into an integrated framework incorporating human resource management, project management, engineering, maintenance planning and enablers such as information and operational technology across the asset lifecycle. The concept of EAM is unique and complex within the business environment [59].

A.G. Kuusk (✉) · J. Gao
University of South Australia, Mawson Lakes, South Australia, Australia
e-mail: anastasia.kuusk@unisa.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_147

The unique and complex nature of EAM is due to the fact that it is different from other business functions such as finance. For example, engineering assets exist as objects independent of contracts between legal entities [2] and are of a specialised nature [29] often producing data as opposed to information, creating unique challenges for organisations with engineering assets. Such specialisation has led to the identification of Operational Technology. Operational Technology (OT) can be found within asset intensive organisations that have hardware or software that detect or cause a change through the direct monitoring and or control of physical devices, processes and events [51]. An example is Supervisory Control and Data Acquisition (SCADA) monitoring and controlling water or energy utility assets. This chapter provides asset infrastructure organisations with recent research results indicating when asset infrastructure organisations should consolidate OT and IT, who should be involved, and how to reduce the OT and IT divide to leverage competitiveness.

2 Understanding the Technology Divide in Engineering Asset Management

Due to the nature of OT it is likely to be implemented and managed by engineering areas of organisations separate to information technology functions [26, 53] and have distinct elements which appear in Table 1. Both technologies share the inherent characteristic of information, although IT is characterised by non-real time decision information and OT by real time asset performance information.

Table 1 The great divide of operational and information technology

Element (Pe = People, Pr = Process, Te = Technology)	Information technology	Operational technology
Budget (Pr)	Dedicated for branch	Embedded within another branches budget
Staff (Pe)	Dedicated IT focus—network analyst	Dual role—engineering and IT maintenance focus
Staff focus (Pe)	Security	Reliability
Objective (Pr)	Strategy-control information	Asset performance-control asset
Systems standards focus (Pr)	COBIT/ITIL	NIST CIP, PAS55, ISA-95
Examples (Te)	Asset management system	SCADA
Information type	Information non real time	Data real time
Networks (Te)	Consolidated	Own network beyond firewall
Uptime (Pr)	Down for patching/backups	100 %

Source Adapted from Gartner model Steenstrup [51]

2.1 The Operational and Information Technology Divide

The divide between the two technologies expose organisations to legal, financial and reputational risks [22, 40, 46] and reduce competitiveness [47]. Increasingly technology is used to control and manage the performance of the assets and provide real time information to asset maintainers, consumers and corporate management. An example of consolidated OT and IT in a power asset infrastructure organisation is provided in Fig. 1.

The consolidation provides for real time dual delivery of consumption information to the household to decide upon consumption, the generator for power provision and the retailer for billing. In the future consolidation provides for next generation asset care by controlling power use within the home and power shedding requirements. Currently the consolidation is impeded by firewalls between the two networks to reduce impacts of perceived security risks of SCADA systems to corporate networks or shutting down of supply by non authorised elements. Numerous benefits of consolidating OT and IT systems, people and process have been identified across the asset management lifecycle. Benefits include reduced hardware and network requirements, integrated life cycle management, reduced licensing costs, data quality improvements from automation, leveraging skilled information technology staff, improving strategic decision making from retrievable holistic single dashboard view of the organisations information and improving competitive use of utility assets [6, 8, 18, 24, 33, 29, 42, 58, 60, 63]. At an engineering operational level Parekh et al. [46] argues the convergence of

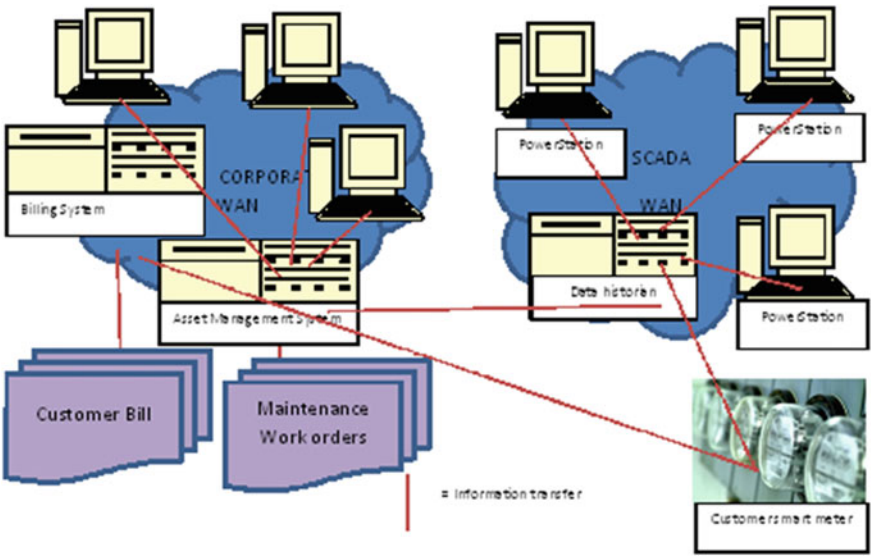


Fig. 1 Example of operational and technology consolidation in a power asset infrastructure organisation

operational and information technologies provides enhanced delivery of utility services. Jaffe et al. [26] indicates the gaps of such a converged environment include funding, technical convergence of technology architectural layers and governance of complex components.

To realise the benefits and facilitate competitive advantage in asset infrastructure organisations it is important to define an accepted, asset infrastructure industry perspective definition of the consolidation taxonomy. Once a baseline definition is established, gaps such as responsibilities, when to move between consolidation phases and key success factors can be established specifically for the asset infrastructure context. How the research was undertaken to study the gaps is described below.

2.2 Reducing the Operational and Information Technology Divide

This research aims to provide asset infrastructure organisations with a framework to reduce the divide between OT and IT. To identify gaps a literature review was undertaken. Success factors for information system leveraging include management support [15, 43, 47, 61, 66], interoperability of platforms and standards [25, 39, 45, 54, 60], enterprise wide asset care and information governance [9, 44, 64, 46, 13, 36] and cross sharing of skills [7, 19, 42, 51]. What is lacking in the literature is the success factor applicability to consolidation of OT and IT in asset infrastructure organisations.

Whilst the literature indicates successful EAM requires consolidation of OT and IT [22, 38, 50, 59] and that financial, reputational and legal drivers exist [16, 48, 51, 65], the literature conflicts as to whether IT or Engineering should be responsible [28, 51], highlights several frameworks, standards and principles that may be applied [20, 23, 25, 53, 58, 46] and that the terminology relating to the phases of consolidation is used interchangeably. The literature also lacks specifics on when organisations move between convergence, integration and alignment of OT and IT.

3 Research Framework

To provide a current practitioners perspective of taxonomy points and success factors for asset infrastructure organisations, three rounds of surveys conforming to Delphi method were undertaken. Chua and Garrett [11], Myers [41] and Kaplan and Maxwell [27] identify the need to choose research methods which allow the study of the social context of practice. The Delphi qualitative research method was chosen to meet the research aims as it facilitates, through consensus of expert practitioner, understanding of information integration in the asset management social context. The method provides for subjective individual judgements, meeting time and cost

efficiencies, a way of efficiently structuring group dialogue and bringing together different organisational functions such as IT and Engineering in a non-competitive environment (Sitt-Ghodes and Crews [58, Powell 56, Turoff 62].

3.1 Research Method

Eight questions relating to why, when and how asset infrastructure organisations converge, align and integrate technology were sent to 30 practitioners between October 2012 and March 2013. Questions appear in Table 3. Theoretical sampling of practitioners as opposed to statistical sampling was used in the study to facilitate validity and reliability through replication as defined by Yin [67], model building and applicability of theory [5, 14]. Experts were defined as Engineering or Information Technologists working in or with asset intensive organisations consolidating operational and information technologies. Fifteen consistent responses from consultants, energy utilities, government asset providers, mining and councils were received between November 2012 and January 2013.

To facilitate the current research meeting the research objectives of credibility, fitting, auditable and confirmable [34] and construct, internal and external validity and replication [67] the Interval Quartile (or interquartile) Range (IQR), mean and standard deviation were calculated as these statistics were used to indicate consensus in original Delphi studies such as Dalkey [12] and Linstone and Turoff [35] in information system contexts. A legend describing the statistics and level of consensus achieved based on the combination of Inter Quartile Range, mean and standard deviation appear in Table 2.

3.2 Research Findings

Practitioner ranked consensus from three rounds of survey conforming to Delphi method is summarised by question in Table 3.

The results provide for asset intensive organisations a framework for consolidating OT and IT to decrease financial, legal and reputational risks and provide competitiveness. The results also provide a current research based taxonomy baseline for consolidation phases applicable to the unique EAM context. The EAM consolidation phases appear in Table 4.

Table 2 Legend of statistics used to identify practitioner consensus

Mean	Standard deviation	Interquartile range	Likert scale
X	Σ	IQR	1 = Never
Strong consensus	Strong consensus	Strong consensus	3 = Sometimes
above 8	less than 2	less than 1	5 = Always

Table 3 Summary of practitioner responses, consensus and ranking of responses to elements of OT and IT convergence, alignment and integration

Question	Strong	Medium	Low	None
1. Convergence undertaken by vendors and alignment and integration activities by organisations?		Vendors converge, organisations align and integrate (66 %, ($\bar{X}$ 1.58, σ 0.92))		
2. Why organisations align and integrate?	Efficient exchange of data; efficient management of information ($\bar{X}$ 4.55, σ 0.52, IQR −1) increased reliability ($\bar{X}$ 4.36, σ 0.67, IQR −1)	Medium decreased costs ($\bar{X}$ 3.55, σ 0.82, IQR −1); single platform ($\bar{X}$ 3.09, σ 0.83, IQR 0)		Increased security
3. How should organisations align and integrate?		Business analysis ($\bar{X}$ 4.36, σ 0.81, IQR −1); joint business effort ($\bar{X}$ 4.27, σ 0.679 IQR −1); standardised platforms ($\bar{X}$ 3.64, σ 0.81, IQR −1)		Vendor involvement
4. What criteria would indicate organisations should move from converging to aligning?		Hardware consistent but applications disparate ($\bar{X}$ 4.09, σ 0.7, IQR −0.05); business needs accounted for ($\bar{X}$ 4.36, σ 0.81, IQR −1); costs ($\bar{X}$ 43.45, σ 0.82, IQR −1)		Only done by vendors
5a. What are the technology success factors for aligning and integrating?			Interoperable solutions ($\bar{X}$ 3.82, σ 0.87, IQR −1.5)	Acceptance of open source solutions
5b. What are the organisation success factors for aligning and integrating?	Agreed enterprise level architecture ($\bar{X}$ 3.91, σ 0.54, IQR 0)	Strategic vision ($\bar{X}$ 4.27, σ 0.47, IQR −0.5); research, plan and execute ($\bar{X}$ 4.45, σ 0.69, IQR −1); open data and communication standards ($\bar{X}$ 4.18, σ 0.6, IQR −0.5); manage as a project ($\bar{X}$ 4.18, σ 0.75, IQR −1); mutual collaboration ($\bar{X}$ 4.09, σ 0.71, IQR −1	Robust framework ($\bar{X}$ 3.91, σ 0.7, IQR −1); systems thinking analysis ($\bar{X}$ 3.91, σ 0.7, IQR −0.5)	

(continued)

Table 3 (continued)

Question	Strong	Medium	Low	None
5c. What are the people success factors for aligning and integrating?		Input from all ($\bar{X}$ 4.27, σ 0.9, IQR −1); ease of use ($\bar{X}$ 3.55, σ 0.82, IQR −1); engineering and IT role to catalyse business change ($\bar{X}$ 3.64, σ 0.81, IQR −1)	Appropriate training ($\bar{X}$ 4.18, σ 0.87, IQR −0.51)	Acknowledge office of CIO
6. What criteria would indicate organisations should move from aligning to integrating?	When data use requires it ($\bar{X}$ 3.82, σ 0.75, IQR 0); when IT/OT structures aligned ($\bar{X}$ 3.82, σ 0.75, IQR 0) and when market competitiveness requires it ($\bar{X}$ 4.36, σ 0.67, IQR −1)	When data use requires it ($\bar{X}$ 3.82, σ 0.75, IQR 0); when IT/OT structures aligned ($\bar{X}$ 3.82, σ 0.75, IQR 0) and when market competitiveness requires it ($\bar{X}$4.36, σ 0.67, IQR −1)		When business and IT have consensus
7. Can information governance facilitate consolidation of Operational and Information Technologies?		Information governance facilitate enterprise level technology change coordination ($\bar{X}$ 4.18, σ 0.75, IQR −1)	Governance informs strategy ($\bar{X}$ 4.09, σ 0.83, IQR −1.5)	
8. IT or Engineering best suited to align or integrate OT and IT?	Combined ($\bar{X}$ 4.73, σ 0.47, IQR −0.5); mutual collaboration ($\bar{X}$ 4.09, σ 0.71, IQR −1			

Table 4 Establishing a baseline OT and IT consolidation taxonomy

New asset infrastructure based taxonomy	Existing Convergence, alignment and integration taxonomies		
Kuusk et al. [30]	Steenstrup [52]	Hoque et al. [21]	Teo and King [57]
Convergence (hardware provided by vendor; consolidated engineering and IT vision of convergence strategy, standards, planning)	Converge (OT and IT share same client, server, network tiers IT and IP based activities often undertaken by vendor)	Alignment (technology supports, enables and not constrains business strategies)	Sequential integration (business goals considered, formulate IS strategy to perform business strategy)
Alignment (architecture aligned by IT and engineering; hardware in place but applications disparate)	Align (after convergence, leading to synchronized standards and architecture plans)	Synchronisation (IS expert resources, support business strategy)	Reciprocal integration (IS expert resources, support business strategy)
Integration (efficient exchange of information and data; driven by market competition and cost savings)	Integrate (outcome of alignment pending bandwidth reduction and firewall conflicts, integrity and reliability of OT and IT)	Convergence (business and technology activities intertwining and leadership teams interchangeable)	Full integration (joint development of strategies, senior management involvement, critical to success of business)

4 Discussion

Lee and Hsu [32], Teo [57], Tarafdar and Ragy-Nathan [55] identify differences in the stages of convergence, alignment and integration, identified by previous researchers as stages of consolidating people, processes and technology for competitive advantage. The main constructs of people, processes and technology govern information for strategic, timely and qualitative decision making, resulting in competitive advantage [31]. Managing information was the most important factor identified in responses to any Delphi study questions in the current research, confirming information as an important element in bridging the OT/IT divide.

The current research also confirmed Steenstrup [51] view that vendors complete convergence activities and organisations align and integrate OT and IT. Consensus on staff involved in aligning and integrating in an organisation indicated a joint effort by Engineering and IT. This contrasts to literature on OT and IT consolidation by engineers [28, 65] arguing engineers are best suited to undertake consolidation activities. Steenstrup [51] earlier research suggest IT standards such as COBIT should be applied to engineering systems with later research [52] indicating a joint approach between IT and engineering is preferred.

Critical success factors for technology such as planning, a holistic enterprise wide view, costs, role involvement and management support [47], Tarafdar and Ragy-Nathan [55] were highlighted in the current research. The current research

differed from the literature in relation to security of OT systems if on IT networks. Since 2011 several conferences, papers and newspaper articles in Australia addressed the issue of security of IT networks if OT is incorporated into them [3, 4, 10, 17, 37, 68]. This coincided with the United States legislation mandating NERC CIP standards for the energy industry in the United States and announcement by the Australian federal government of a Cyber Security Centre. Consensus was not reached on this being a factor for why organisations consolidate OT and IT in the current survey results.

The survey results from a practitioner perspective indicates asset infrastructure organisations should;

1. Plan for convergence when external factors such as a corporate vision and consolidated industry standards are in place. Organisations should prepare by analysing business needs and objectives, planning and research options available and developing a convergence strategy. At this point vendors may sell a vision of convergence to organisation.
2. Move to convergence when there is consensus between business and IT. Convergence is established when vendors provide hardware which is IP addressable and has the same chips and routers as provided in other parts of the organisation and engineering, information management and IT provide input into application development.
3. Move from convergence to alignment when the hardware is in place but applications and information are disparate. Alignment is characterised by an architecture aligned by IT and Engineering with advice provided by vendors.
4. Move from alignment to integration when market competition and need for cost savings arise. The integration stage is characterised by enterprise wide data exchange.

5 Conclusion

This chapter provides for asset intensive organisations, vendors and consultants what the consolidation of OT and IT phases are as a new context specific taxonomy. The taxonomy identifies when organisations should converge, align and integrate OT and IT and whom should be responsible, from practitioners' perspectives. The chapter identifies how asset intensive organisations, by consolidating people, process and technology, can bridge the great wall of Operational and Information Technologies for competitive advantage. Further qualitative case study research will be undertaken to explore particular topics raised by the research so far, such as what engineering standards asset infrastructure organisations are utilising for OT and IT convergence, how organisations overcome the IT security versus engineering reliability and uptime focus and if information governance can facilitate OT and IT convergence, alignment and integration.

References

1. Australian Asset Management Collaborative Group (2011) Guide to integrated strategic asset. Australian Asset Management Collaborative Group, Brisbane
2. Amadi-Echendu J (2010) What is engineering asset management? In: Amadi-Echendu J, Brown K, Willett R, Mathew J (eds) Definitions, concepts and scope of engineering asset management. Springer, London
3. Barwick H (2012) SCADA systems in Australia easy target for malware: security expert. CIO Australia. 9 August 2012
4. Beggs C (2008) A holistic SCAD security standard for the Australian context. In: Proceedings of the 9th Australian information warfare and security conference. Edith Cowan University, Perth Western
5. Benbasat I, Goldstein D, Mead M (1987) The case research strategy in studies of information systems. MIS Q 11(3):369–386
6. Berst J (2011) Smart grid technologies putting pressure on utility IT spending, says new report. SmartGrid News. 26 Oct 2011
7. Boone T, Ganeshan R (2008) Knowledge acquisition and transfer among engineers: effects of network structure. Manag Decis Econ 29(5):459–468
8. Bonnet P (2010) The sustainable it architecture: resilient information systems. Wiley, Sydney
9. Caldwell F (2011) Hype cycle for legal and regulatory information governance. Gartner, USA
10. Chaudary (2012) Is the business network connected to SCADA? Need for Auditing SCADA Networks. ISACA J 6:1–8
11. Chuachia F, Garret T (2009) Understanding ontology and epistemology in information system research. IGI Global, Hershey
12. Dalkey NC (1969) An experimental study of group opinion. Futures 1(5):408–426
13. DeuBois L, Tero V (2010) Practical Information governance: balancing cost, risk, and productivity. IDC, Framingham
14. Eisenhardt K (1989) Building theories from case study research. Acad Manag Rev 14(4):532
15. Evans S (2010) CBR Data Governance Survey 2010. Computer Business Review, London
16. Fishwick P (1996) Toward a convergence of systems and software engineering. Int J Gen Syst 17(1):1–20
17. Griffith C (2012) Hackers tap into local essential services. The Australian, 7 August
18. Haider A (2010) Enterprise architectures for information and operational technologies for asset management. In: Mathew J, Ma L, Tan A, Weijnen M, Lee J (eds) Engineering asset management and infrastructure sustainability: 5th world congress on engineering asset management, Brisbane, Australia, p 315
19. Haider A (2011) IT enabled engineering asset management: a governance perspective. J Org Knowl Manag. doi: 10.5171/2011.348417
20. Hillard R (2010) Information-driven business: how to manage data and information for maximum advantage. Wiley, New Jersey
21. Hoque F, Trainer T, Wilson C (2005) Winning the 3-Legged race:when business and technology run together. prentice Hall, New Jersy
22. Humphrey B (2003) Asset Management, in theory and practice. Energy pulse: insight, analysis and commentary on the global power industry 50–53
23. IBM (2007) Creating a competitive advantage with converged communications. IBM Global Services, USA
24. Institute of Public Works Engineering Australia (2011) International infrastructure management manual. Institute of Public Works Engineering Australia, Sydney
25. International Electrotechnical Commission (2007) Application integration at electric utilities: system interfaces for distribution management, 61968. International Electrotechnical Commission, Geneva
26. Jaffe S, Torchia M, Feblowitz J, Nicholson R (2011) The virtual power plant: integrating operations technology and IT. IDC Energy Insights, Framingham

27. Kaplan B, Maxwell J (1994) Qualitative research methods for evaluating computer information systems. In:Evaluting health care information systems: Methods and applications. SAGE, Thousand Oaks
28. Kern A (2009) IT/automation convergence revisited: keeping automation close-coupled to operation is key. Hydrocarbon Process 88:61–62
29. Koronios A, Steenstrup C, Haider A (2009) Information and operational technologies nexus for asset lifecycle management. In: Kiritsis D (ed) 4th world congress on engineering asset management, Athens, Greece, p 112
30. Kuusk A, Koronis A, Gao J(2013) overcoming integration challenges in organisations with operational technology. In: Proceedings of the 24th Australian conference on information systems (Aus) melbourne, Australia
31. Leavitt H (1965) Applied organizational change in industry. Carnegie Institute of Technology, Graduate School of Industrial Administration, Pittsburgh
32. Lee, Hsu (2009) "The evolution of planning for information systems. In: King WR (ed) Planning for information systems. Advances in management information systems, vol 14. NY, M.E. Sharpe, pp 318–340
33. Lin S, Gao J, Koronios A, Chanana V (2007) Developing a data quality framework for asset management in engineering organisations. Int J Inf Qual 1(1):100–126
34. Lincoln Y, Guba E (1985) Naturalistic inquiry. Sage, Beverly Hills
35. Linstone HA, Turoff M (2002) The Delphi method: techniques and applications. Addison-Wesley Publishing Company, Boston (reading)
36. Logan D (2010) Cooperation is key for managing e-discovery. Gartner, USA
37. Mahoney J, Iyengar P, Roberts J, Steenstrup K (2013) Growth of IT/OT convergence creates risks of enterprise failure. Gartner, USA
38. May Business School (2011) Texas A&M asset performance management study. Texas A&M University, College Station
39. McDonnell Group (2012) Integration of IT and OT: enabling the electric grid of the future. The McDonnell Group, California
40. McManus J (2004) Information governance: an ethical perspective. Manag Serv 48(12):16–17
41. Myers MD (1997) Qualitive research in information systems. MIS Q 21(2):241–242
42. Newman D (2011) Enterprise architecture research index: how enterprise information architecture improves information sharing. Gartner, USA
43. Nfuka E, Rusu L (2011) The effect of critical success factors on IT governance performance. Ind Manag Data Syst 111(9):1418–1448
44. Nicolett M, Proctor P (2011) MarketScope for IT governance, risk and compliance management. Gartner, USA
45. Office of the National Coordinator for Smart Grid Interoperability (2010) NIST framework and roadmap for smart grid interoperability standards. National Institute of Standards and Technology, Gaithersburg
46. Parekh K, Zhao J, McNair K, Robinson G (2007) Utility enterprise information governance strategies (2007). Grid Interop Forum, Albuquerque
47. Rockart J (1979) Chief executives define their own data needs. Harvard Bus Rev 57(2):81–93
48. Romero C (2011) Convergence of information and operation technologies (IT & OT) to build a successful smart grid. Ventyx, Atlana
49. Sitt-Ghodes W, Grews T (2004) Delphi technique: A research strategy for career and technical education. J Career Tech Educ 20(2):55–67
50. Sklar D (2004) Principles of asset management: the holistic model. In: Energy pulse: insight, analysis and commentary on the global power industry. Energy Central, Aurora
51. Steenstrup K (2008) IT and OT: intersection and collaboration. Gartner, USA
52. Steenstrup K (2010) Operational technology convergence with IT: definitions. Gartner, USA
53. Steenstrup K (2011) IT and operational technology: convergence, alignment and integration. Gartner, USA
54. Steenstrup K, Johnson G, Mahoney J, Perkins E (2012) Agenda for OT/OT alignment. Gartner, USA

55. Tarafdar M, Ragu-Nathan T (2012) Business information system alignment: taking stock and looking ahead. Planning for information systems. Benchmarking. Int J 19(4–5):604–617
56. Taylor-Powell E, Renner M (2003) Analyzing qualitative data. University of Wisconsin Extension, Madison
57. Teo T, King W (1997). Integration between business planning and information systems planning: An evolutionary-contingency perspective. J manag Inf syst 14:185–214
58. Thomas G (2009) The DGI data governance framework. Data Governance Institute, Orlando
59. Too E (2010) A framework for strategic infrastructure asset management. In: Amadi-Echendu J, Brown K, Willett R, Mathew J (eds) Definitions, concepts and scope of engineering asset management. Springer, London
60. Torchia M (2011). Network Interoperability Is Key to Success. IEEE Smartgrid Newsletter, May
61. Trkman P (2010) The critical success factors of business process management. Int J Inf Manage 30:125–134
62. Turoff M (1970) The design of a policy delphi. Technologies forecasting and soical change 2 (2)
63. Waddington D (2008) Adoption of data governance by business. DM Rev 18(12):32–34
64. White McManus J Atherton A (2007) Governance and information governance: some ethical considerations within an expanding information society. Int J Qual Stan 1(1):180–192
65. Wiese I (2004) The integration of SCADA and corporate IT. In: Australian SCADA conference
66. Yeoh W, Gao J, Koronios A (2009) Empirical investigation of critical success factors for implementing business intelligence systems in multiple engineering systems in multiple engineering asset management organisations. In: Cater-Steele A, Al hakim L (eds) Infrastructure system research methods, epistemology and applications. Information Science, New York
67. Yin RK (2009) Case study research, design and methods. Sage Publications, Newbury Park
68. Zimmerman R, Fraser J (2007) Come together: IT controls engineering convergence furthers manufacturers' success. Rockwell Automotation, Milwaukee

Calculation of the Expected Number of Failures for a Repairable Asset System

Gang Xie, Lawrence Buckingham, Michael Cholette and Lin Ma

Abstract The expected number of failures is the essential element in cost analysis for a repairable system in engineering asset management. A renewal process is typically used for modelling a repairable system with perfect repairs while a non-homogeneous Poisson process can be used to model a repairable system with minimal repair. An asset system with imperfect repair will be restored to the state which is somewhere between as bad as old and as good as new. While imperfect repairs are more realistic, it is more challenging to calculate the expected number of failures. In this chapter, we propose an imperfect reparable system assuming decreasing restoration levels conditional on the previous repair actions. Compared with a popular imperfect repairable system settings which assumes a constant discount restoration level after the first failure occurrence, our decreasing restoration levels model may better represent the actual repair-restoration patterns for many asset systems. The likelihood function of the newly proposed model is derived and the model parameters can be estimated based on historical failure time data. We adopt a cumulative hazard function based Monte Carlo simulation approach to calculate the expected number of failures for the newly proposed reparable system model. This new simulation algorithm is demonstrated on both simulated and real data and compared to a popular existing model under a Weibull distribution setting. An advantage of our simulation algorithm is that a bootstrap version confidence band on the estimated expected number of failures can easily be constructed. The modelling and simulation results in the chapter can be used for the development of an engineering reliability analysis and asset management decision making tool.

G. Xie (✉) · L. Buckingham · M. Cholette · L. Ma
Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia
e-mail: john.xie@qut.edu.au

L. Buckingham
e-mail: l.buckingham@qut.edu.au

M. Cholette
e-mail: michael.cholette@qut.edu.au

L. Ma
e-mail: l.ma@qut.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_148

Keywords Expected number of failures · Imperfect repair · Virtual age · A decreasing restoration levels model · Cumulative hazard function · Monte Carlo simulation · Bootstrap confidence band

1 Introduction

A crucial step in determining maintenance policies and asset life costs is the computation of the expected number of failures for an asset. Renewal processes (RPs) are typically used for modelling repairable systems with perfect repairs (i.e. system is restored to *as good as new* after each repair). Apart from the homogeneous Poisson processes (HPP), a closed form solution to calculate the expected number of failures is generally not available for RPs [7, 10]. An alternative is to assume repairs are minimal (i.e. *as bad as old*). In such cases, a nonhomogeneous Poisson process (NHPP) can be used to model a repairable system and the expected number of failures can be calculated analytically based on the cumulative hazard function of the point process to the first failure [2, 12, 14, 16]. More generally, an asset system with imperfect repair will be restored to a state which is somewhere between *as bad as old* and *as good as new* [2, 12, 16]. However, due to the stochastic process characteristics, it is challenging to calculate the expected number of failures for an asset system with imperfect repairs [10, 16].

Point process models are the natural and suitable tools for analysis of a repairable asset system because the failure times and the corresponding repair actions form a sequence of recurrent stochastic events. Cox's seminal paper [5] is considered as one of the first comprehensive treatment of statistical methods for recurrent events in the context of engineering reliability analysis and decision making. Many results from [5] are contained in the subsequent book by Cox and Lewis [6]. The modelling and analysis of a repairable system is far more difficult than that of a non-repairable system [12, 13]. For many years, Ascher and Feingold [2] (published in 1984) is the first and only book devoted solely to repairable systems reliability. A more recent text (published in 2000) on reliability of repairable systems is [14] by Rigdon and Basu. Although this book contains intensive mathematical contents for various statistical models in its methodology chapters, the subsequent chapters are more applied supported with S-PLUS or SAS programs for some example questions. Pham and Wang authored a highly cited review paper on imperfect maintenance/repair of repairable systems up to 1996 [13]. Lindqvist [12] advanced the review period up to 2006 on the statistical modelling and analysis of repairable systems. In his review paper, Lindqvist presented a three-dimensional model to summarise a general repairable system model framework based on (1) hazard rate (HPP or NHPP); (2) between failure time interval distribution (HPP or RP); (3) for multi-failure-mode system, the heterogeneity between different failure modes. The popular virtual age based imperfect repair system model first proposed by Kijima [11] is a single-failure-mode model which is equivalent to the Trend Renewal Process (i.e. a generalised RP)

in Lindqvist's 3-D model cube framework. The well-known Brown–Proschan model [3] can be treated as a special case of Kijima's models [12]. In their study in non-parametric statistical inference, Dorado et al. [8] proposed a model slightly more general than Kijima's models. Yañez et al. [16] proposed a counting simulation algorithm to calculate the expected number of failures based on a virtual age based repairable system model which is essentially a special case of Kijima's Model I.

In Sect. 2.1, we give a more detailed description on Kijima's virtual age repairable system models. A new virtual age repairable system model is proposed and the model likelihood function is derived in Sect. 2.2. The different virtual age models are compared and the results are presented in Sect. 2.3. In Sect. 3, we propose a cumulative hazard function based simulation algorithm for calculating the expected number of failures for our newly proposed virtual age model. The new simulation algorithm is verified using a simulated data set and a real data set adopted from [16].

We will assume the repair times are negligible (i.e. the system is repaired and put into new operation immediately after the failure) and a single failure mode in our model specification and data analysis in this chapter. The statistical analysis and simulation sample generation have been performed using R, an open source professional statistical language/package [15]. The R code for implementing the cumulative hazard function based simulation algorithm is presented in Appendix 1 for readers' reference.

2 Virtual Age and the Single Mode Generalised Repairable System Models

Let $f(t)$ be the probability density function (pdf) and $F(t)$ the cumulative distribution function (cdf) for the time to the first failure of a repairable asset system. Hence, by definition, the reliability function is $R(t) = 1 - F(t)$, the hazard function is $h(t) = f(t)/R(t)$ and the cumulative hazard function is $H(t) = \int_0^t h(u)du$ for $t > 0$.. Data from repairable systems are usually recorded as ordered failure times $t_1, t_2, \ldots$ and we denote the time intervals between failures by $X_1, X_2, \ldots$ accordingly as shown in Figs. 1 and 2. Further, let $N(T)$ be the number of failures over the time period $(0, T]$ and $E(\cdot)$ be expectation. Therefore by definition, $H(t_1) = H(X_1) = E[N(t_1)]$, i.e. $H(t_1)$ is the expected number of failures over the time period $(0, t_1]$.

2.1 Definition and Interpretation of Asset System Virtual Ages

Kijima [11] first proposed the virtual age concept in construction of a statistical model for repairable systems. The 'virtual age' is determined by the restoration level achieved after each failure/repair on the asset system. More specifically, a

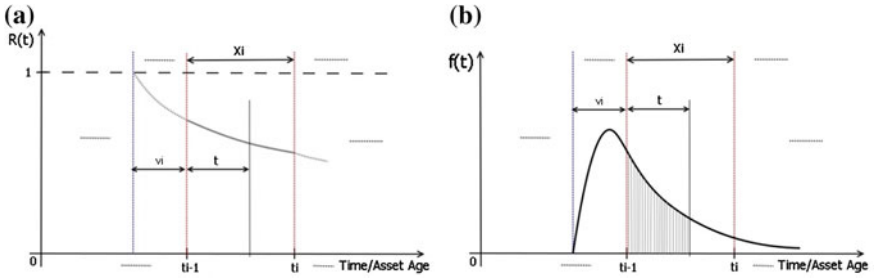


Fig. 1 Definition of asset virtual age in terms of reliability function $R(t)$ and probability density function $f(t)$ where X_i is the time interval between the $(i-1)$th and ith failures

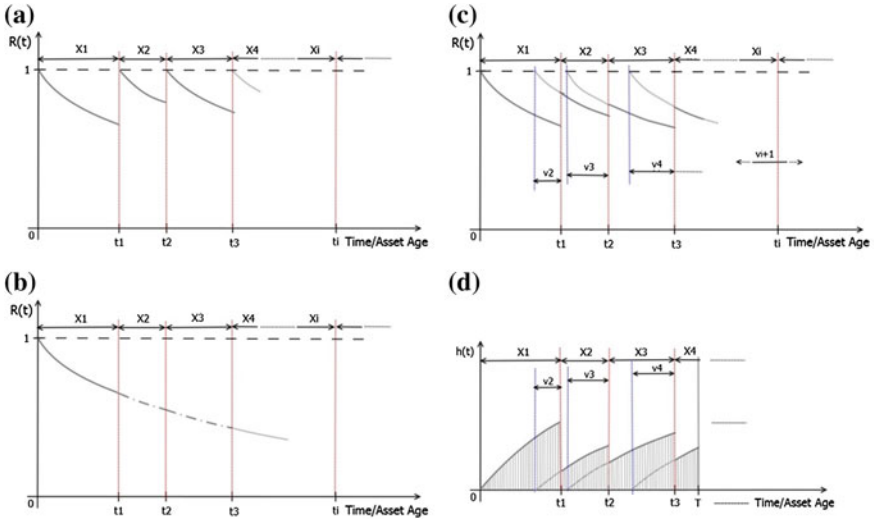


Fig. 2 Asset virtual age, repairable system models, and calculation of the expected number of failures

system with virtual age $v \geq 0$ is assumed to behave as if it were a new system which has reached age v without having failed. This definition is schematically illustrated in Fig. 1.

2.2 Model Specification and Parameter Estimation

Let v_i denote the virtual age of an asset system at the ith failure time t_i. We propose a new virtual age based repairable system model labelled as Model A:

$$v_1 = 0;\ v_i = \left(1 - q^{i-1}\right)t_i = \left(1 - q^{i-1}\right)\sum_{j=1}^{i-1} X_j \quad \text{for } i > 0 \text{ and } j < i.$$

Yañez et al. [16] proposed a virtual age based model as:

$$v_1 = 0;\ v_i = qt_i = \sum_{j=1}^{i-1} X_j \quad \text{for } i > 0 \text{ and } j < i. \text{ This is termed Model B.}$$

Note that when q = 1, Model A is a renewal process with perfect repair, and when q = 0, it becomes a NHPP model with minimal repair. For Model B, q = 0 corresponds to perfect repair process while q = 1 models minimum repair. Thus Models A and B both represent generalised repairable system models albeit with differing parameterisations.

We note that Kijima's Model I is specified as $v_1 = 0;\ v_i = Q_i t_i = Q_i \sum_{j=1}^{i-1} X_j$ for $i > 0$ and $j < i$ where Q_i is a random variable on the interval $\lfloor 0, 1 \rfloor$. If $Q_i = q$ (a constant), Model I reduces to Model B. Kijima's Model II is specified as $v_1 = 0;\ v_i = \sum_{j=2}^{i} \left(X_j \prod_{k=j}^{i} Q_k\right)$ for where once again Q_k is a random variable. If $Q_k = q$ (a constant), Model II reduces to $v_i = \sum_{j=2}^{i} q^{i-j+1} X_j$. It is trivial to show that given the same $q \in \{0, 1\}$ and observed failure times, virtual ages calculated via Model II are bounded above by the virtual ages calculated via Model I.

The perfect repair RP model, the minimal repair NHPP model, and the virtual age based imperfect repair model are schematically illustrated in Figs. 2a–c, respectively.

Note that Model B can be considered to be a constant discount restoration level (towards the *as bad as old* level) model. Model A is a decreasing restoration level model which is not a special case of Kijima's models. With a certain $q > 0$ value, Model B will never reach the *as bad as old* level. On the other hand, with Model A, at some point as the number of failures increases, the system will eventually reach a state in which repairs are effectively *as bad as old*. This may better represent the actual repair-restoration patterns for many asset systems. This fundamental difference is clearly demonstrated by a simulation analysis for comparing the calculated virtual ages by Model A and Model B.

Table 1 gives the simulation sample data details and Table 2 presents the virtual age comparison results.

It is well known that the likelihood function of a virtual age based repairable system model can be derived based on the conditional probability formula (referring to the shaded area in Fig. 1b)

$$\Pr[X_i \le t | v_i] = \frac{F(t + v_i) - F(v_i)}{1 - F(v_i)} \tag{1}$$

Table 1 Simulated failure time sample (β = 2.5, η = 300, q = 0.85, Model A, rounded to integers)

Index (i)	1	2	3	4	5	6	7	8
Time between failures (X_i)	299	429	10	57	101	19	46	33
Failure times (t_i)	299	728	738	795	896	915	961	994

Table 2 Model comparison based on the simulated data

Virtual age	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
Model A: $\hat{\beta} = 3.10, \hat{\eta} = 460,$ $\hat{q} = 0.536, AIC = 89.3$	139	519	624	729	856	893	949	987
Model B: $\hat{\beta} = 4.37, \hat{\eta} = 498,$ $\hat{q} = 0.705, AIC = 89.9$	211	513	520	560	632	645	677	701

where $F(t)$ is the cdf for the first time to failure of the system as defined in the beginning of Sect. 2 [11–13, 16]. In this chapter, we choose the standard two-parameter Weibull distribution to be the interval representation of a renewal process for model specification and subsequent data analysis. Hence, the first failure time t_1 has pdf

$$f(t) = \left(\frac{\beta}{\eta^{\beta}}\right) t^{\beta-1} \exp[-\left(\frac{t}{\eta}\right)^{\beta}]. \tag{2}$$

Without loss of generality, we consider the failure terminated case, so that the sample likelihood function is

$$\begin{aligned} L(t_n) &= f(t_1, t_2, \ldots, t_n) \\ &= f(t_1) f(t_2|t_1) f(t_3|t_2, t_1) \ldots f(t_n|t_{n-1}, t_{n-2} \ldots t_1) \\ &= f(t_1) f(t_2|t_1) f(t_3|t_2) \ldots f(t_n|t_{n-1}) \end{aligned} \tag{3}$$

Combining Eqs. (1) and (2) yields the pdf of the virtual age based repairable system as

$$f(t|t_i) = \left(\frac{\beta}{\eta^{\beta}}\right) [t + v_i]^{\beta-1} \exp[\left(\frac{v_i}{\eta}\right)^{\beta} - \left(\frac{t + v_i}{\eta}\right)^{\beta}] \tag{4}$$

for $i = 1, 2, \ldots, n$ (n is the sample size). By Eqs. (3) and (4), the sample likelihood function for Models A and B can be derived as

$$L(t_n) = \left(\frac{\beta}{\eta^{\beta}}\right)^{n} (t_1)^{\beta-1} \exp[-\left(\frac{t_1}{\eta}\right)^{\beta}] \prod_{i=2}^{n} \left\{ [t_i + v_i]^{\beta-1} \exp[\left(\frac{v_i}{\eta}\right)^{\beta} - \left(\frac{t_i + v_i}{\eta}\right)^{\beta}] \right\} \tag{5}$$

2.3 Model Comparison

Based on Eq. (5), the model parameter estimation can be done by following the standard Maximum Likelihood Estimation (MLE) procedure. The difference between Model A and Model B in applying Eq. (5) for parameter estimation is in the different specification of virtual age v_i. In this chapter, the built-in R function 'optim' is used for performing MLE procedure. Table 3 gives a real failure time data set which is adopted from a case study undertaken on USS Halfbeak [16]. Table 4 presents the results obtained by fitting each candidate model to this dataset. Model RP stands for Renewal Process model, i.e. $q = 1$ in Model A setting or $q = 0$ in Model B. Goodness-of-fit is measured by *Akaiki Information Criterion* (AIC) [1] with a lower score indicating a better fit to the observed data. A difference in AIC scores of less than two units is not considered to be statistically significant [1, 4]. Although the parameter estimates obtained for Models A and B differ markedly due to the different parameterisation of virtual age, the AIC scores are not significantly different. In this respect, Models A and B are equally effective while Model RP is marginally weaker. This implies that we should not assume an *as good as new* repairable system for the USS Halfbeak example data.

3 Calculation of the Expected Number of Failures

This section sets out a method by which the expected number of failures may be computed for a repairable system subject to imperfect repair.

3.1 A Brief Review on the Existing Approach

A closed form solution to calculate the expected number of failures $E\lfloor N(T)\rfloor$ is generally not available for renewal processes [7, 10]. However, Jardine [10 pp. 44–46] presents an iterative numerical algorithm to calculate an approximate value of $E\lfloor N(T)\rfloor$ for such a system. On the other hand, a closed form solution is available for a system with minimal repairs based on the cumulative hazard function of the first failure time process: $E[N(T)] = \int_0^T h(u)\,du$ [2, 12, 14, 16]. Neither of these methods applies to a system with imperfect repairs, leaving Monte Carlo simulation as the only viable option. The most intuitive and direct way is to generate a large number of simulated sample datasets, count the realised failures in each simulated sample dataset and then average to obtain an estimate. Yañez et al. [16] proposed one of such counting simulation algorithm. Based on the fact that, for the first failure time t_1, $E\lfloor N(T = t_1)\rfloor = H(t_1)$, we propose a simulation algorithm based on the cumulative hazard to calculate the expected number of failures following the approach outlined in Sect. 3.2. In the comparative study detailed below

Table 3 Failure time data adopted from Table 3 in [16] (USS Halfbeak example, in hours)

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
860	2,468	3,602	6,305	6,950	7,045	8,323	8,928	9,272	10,326	11,006	11,411
X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	X_{21}	X_{22}	X_{23}	X_{24}
11,778	14,536	14,891	15,975	16,830	17,110	17,600	18,545	18,650	18,777	18,838	19,164

Table 4 Model comparison based on the USS Halfbeak example dataset

Model A	$\hat{\beta}$ = 1.778, $\hat{\eta}$ = 1,535, $\hat{q}$ = 0.9899, AIC = 368.0
Model B	$\hat{\beta}$ = 2.054, $\hat{\eta}$ = 1,873, $\hat{q}$ = 0.1575, AIC = 368.5
Model RP	$\hat{\beta}$ = 1.177, $\hat{\eta}$ = 8,47.0, q = 1.000, AIC = 371.8

the proposed algorithm is referred to as *Method A* while we refer to the algorithm of [16] as *Method B*. In addition to the simulation methods, the analytic result for the minimal repair case provides an upper bound on the expected number of failures while the iterative method outlined by Jardine for a renewal process provides a lower bound, both of which can be used to check the results obtained via simulation.

3.2 The Cumulative Hazard Function Based Simulation Approach

The cumulative hazard function based simulation algorithm works by generating a large number of Monte Carlo simulation sample datasets representing the operational life of a single asset. From each sample dataset, one $E\lfloor N(T)\rfloor$ value is calculated. More precisely,

$$\begin{aligned} E[N(T)] &= \sum_{i=1}^{n}\left[\int_{v_i}^{v_i+t_i} h(u)\,du\right] \\ &= \sum_{i=1}^{n}[H(v_i+t_i)-H(v_i)] \\ &= \sum_{i=1}^{n}\left[\left(\frac{v_i+t_i}{\eta}\right)^{\beta}-\left(\frac{v_i}{\eta}\right)^{\beta}\right] \end{aligned} \tag{6}$$

for $T = t_n$ where n is the sample size. For example, the shaded area in Fig. 2d can be calculated as $E[N(T)] = \sum_{i=1}^{3}\left[\left(\frac{v_i+t_i}{\eta}\right)^{\beta}-\left(\frac{v_i}{\eta}\right)^{\beta}\right]+\left(\frac{T}{\eta}\right)^{\beta}-\left(\frac{v_A}{\eta}\right)^{\beta}$. The collection of simulated $E\lfloor N(T)\rfloor$ values can be used to construct a bootstrap confidence interval [9]. In this chapter, we use the 2.5 and 97.5 percentile values to form an approximate 95 % confidence interval for $E\lfloor N(T)\rfloor$. Results obtained via the two simulation based methods are presented in the next subsection.

Table 5 Comparison of different methods of calculating the expected number of failures E[N(T)]

Time periods (0, T]		50	100	150	200	250	300
Numeric approximation		0.450	1.114	1.806	2.498	3.190	3.882
Method B		0.451	1.118	1.812	2.508	3.196	3.886
Method A	Mean	0.453	1.120	1.803	2.493	3.171	3.899
	Lower	0.323	0.765	1.306	1.760	2.329	2.921
	Upper	0.494	1.398	2.567	3.789	4.270	5.380
Minimal repair		0.494	1.398	2.567	3.953	5.524	7.262

3.3 Comparison of Different Approaches

Table 5 presents the $E\lfloor N(T)\rfloor$ simulation results for an assumed renewal process with Weibull distribution: $\boldsymbol{\beta} = 1.5$, $\boldsymbol{\eta} = 80$. A modified version of Jardine's algorithm [10, pp. 44–46] is labelled "Numeric approximation" in Table 5. We applied a rule of thumb to determine the number of iteration intervals for this algorithm to ensure an accurate approximation. The number of iteration intervals is taken to be 300 times the highest $h(t)$ value over the period of the simulation. There is a trade-off between approximation accuracy and computation time. An approximate 95 % confidence interval for $E\lfloor N(T)\rfloor$ is computed for Method A and shown in Table 5. One might similarly obtain a bootstrap confidence band for Method B by iterating the simulation step, however this would incur a very high computational cost.

The results in Table 5 should be read as follows. In the Numeric approximation row, 0.450 in the column headed 50 means $E\lfloor N(50)\rfloor \approx 0.450$ as calculated via this method; likewise, $E\lfloor N(100)\rfloor \approx 1.114$ and so forth. The last row gives the expected number of failures under the *as bad as old* assumption. These are upper bounds on the estimated estimates in the corresponding column. Given all the above discussed technical details, Table 5 shows that all these three algorithms produce equally good estimation of $E\lfloor N(T)\rfloor$ using our R code programs developed for this study.

Table 6 shows the results obtained by applying Methods A and B to the models presented in Table 4 which were derived from the the USS Halfbeak example dataset (Table 3). Recall that Model A is our newly proposed model and Model B is that of Yañez et al. [16]. Both models are specified in Sect. 2.2. To illustrate the difference between Methods A and B we apply both approaches to the example data of Table 3 to produce the series labelled "MethodA (observed)" and "MethodB (observed)". Method A calculates $E\lfloor N(T)\rfloor$ directly based on the observed sample data whereas Method B can only give the observed failure counts. A graphical representation of the data from Table 6 is shown in Fig. 3.

Table 6 Calculation of the expected number of failures E[N(T)] (USS Halfbeak case study)

Time periods		0	2,000	4,000	6,000	8,000	10,000	12,000	14,000	16,000	18,000	20,000
MethodA Model A	Lower	0	0.70	1.80	3.25	4.94	6.99	9.51	12.35	15.67	19.38	23.60
	Upper	0	1.60	4.06	6.15	8.08	10.35	12.51	15.22	18.90	23.77	28.60
	Mean	0	1.15	2.66	4.30	6.11	8.18	10.61	13.48	17.00	21.25	25.90
MethodA observed		0	0.95	2.21	4.46	6.10	7.85	10.07	13.61	17.37	21.23	26.20
MethodB observed		0	1.00	3.00	3.00	6.00	9.00	13.00	13.00	16.00	19.00	24.00
MethodB Model B		0	0.88	2.38	4.15	6.21	8.59	11.28	14.28	17.66	21.38	25.47

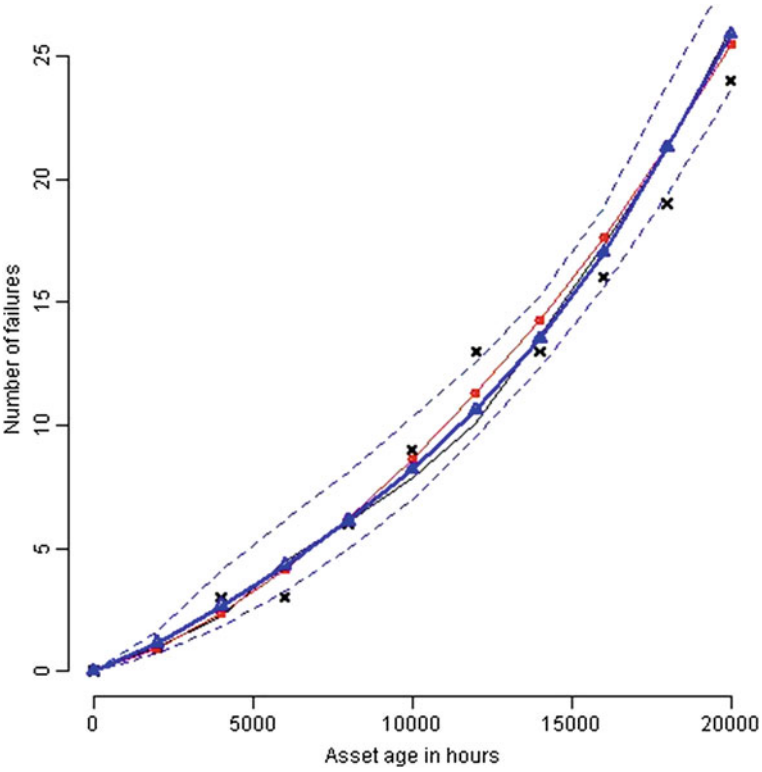


Fig. 3 Verification and comparison of the estimation of expected number of failures (USS Halfbeak's example data). *Blue bold lines* are the expected number of failures calculated by Method A (model A); two *dashed blue lines* are the 95 % confidence band. *Red solid lines* are the expected number of failures calculated by Method B (model B). *Black points* (*crosses*) are the observed number of failures

4 Conclusion

In this chapter, we have proposed a decreasing restoration level virtual age type repairable system model. Different from the existing virtual age type models, as the number of failures increasing, at some point, the system will reach a level that repair action become 'useless', i.e. the restoration level is as bad as old. The system model likelihood function is derived and model parameters can be estimated based on the observed failure time sample data. The model comparison results show that the proposed virtual age type repairable system model can produce a statistically better goodness-of-fit in analysis of real failure time data.

We have examined three different approaches for calculating the expected number of failures for a general repairable system verified on both simulated and real data sets of single type failure mode. The newly proposed cumulative hazard function based simulation algorithm produces as good as expected number of failures results as the existing counting simulation algorithm. The results obtained from simulation algorithms match the results by the iterative numerical approximation algorithm almost perfectly. The cumulative hazard function based simulation approach has the advantage in easy construction of a bootstrap confidence band for the estimated expected number of failures.

The results achieved in this chapter will be used for the development of a reliability engineering analysis and asset management decision making tool. Although the Weibull distribution is chosen for model specification and algorithm verification in this chapter, it is possible to adapt these algorithms to other distribution model specifications. The calculation of the expected number of failures for multi-failure mode repairable systems will be our future research topic.

Appendix

The core R code for calculating the expected number of failures using the cumulative hazard function simulation algorithm proposed in this chapter is included below for readers' reference.

```
#-----------------------------------------------------------------

SIMUCUMUH <- FUNCTION(QFTOR=QFACTOR, BETA1=BETA, ETA1=ETA,

        ST = STOPT, N1 = ITN1) {

         SAMPH = NULL

         FOR(K IN 1:N1){

                 SAMP = SAMPVA =NULL; CUMUT = CUMUH = 0; M = 0

                 WHILE(CUMUT<= ST) {

MIDT = RWEIBULL(1,SHAPE=BETA1,SCALE=ETA1)

IF(M == 0 &&MIDT>=ST) {

                 CUMUH = (ST/ETA1)^BETA1
```

```
                        CUMUT = MIDT + 0.1 }

                        ELSE {    # START OF OUTER 'ELSE'

                        IF(M == 0) {MIDT1 = MIDT; SAMPVA =C(SAMPVA,0)}

                                ELSE {VAGE = (1-QFTOR^M) * CUMUT;

                                  MIDT1 = MIDT-VAGE; SAMPVA =C(SAMPVA,VAGE) }

                        IF(MIDT1 > 0) {SAMP = C(SAMP,MIDT1); CUMUT = CUMUT +
        MIDT1 ; M = M + 1 }

                                } # END OF OUTER 'ELSE'

                                        } # END OF 'WHILE'

        IF(M > 0) {

                        NT = LENGTH(SAMP) -1; SAMP1 = C(SAMP[1:NT],(ST-
        SUM(SAMP[1:NT])))

        #

                        IF(QFTOR==1) SVAGE = SAMPVA

                        IF(QFTOR!=1) SVAGE = UNIQUE(SAMPVA)

                                FOR(I IN 1:LENGTH(SAMP1)) {

                        MIDH = ((SAMP1[I]+SVAGE[I])/ETA1)^BETA1 -
        (SVAGE[I]/ETA1)^BETA1

                        CUMUH = CUMUH + MIDH   }

                           } # END OF 'IF(M > 0)'

                        SAMPH = C(SAMPH, CUMUH)

                                 }

                        MIDBD = AS.NUMERIC(QUANTILE(SAMPH,C(0.025,0.5,0.975)))

                        LOWL = MIDBD[7]; MEDL = MIDBD[10]; UPL =
        MIDBD[2];MEANL = MEAN(SAMPH)

        #

                invisible(list(lowL=lowL,medL=medL,upL=upL,meanL=meanL))

                                 } # end of function
#-----------------------------------------------------------------

# Example

qfactor = 0.95; itN1 = 1000; beta = 1.5; eta = 80; stopT = 300

        expectedfail = simuCumuH(sT=stopT); expectedfail

#-----------------------------------------------------------------
```

References

1. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control AC-19(6):716–723
2. Ascher H, Feingold HC (1984) Repairable system reliability—modelling, inference, misconceptions and their causes. Marcel Dekker, New York
3. Brown M, Proschan F (1983) Imperfect repair. J Appl Probab 20:851–859
4. Burnham K, Anderson D (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer, Berlin
5. Cox DR (1955) Some statistical methods connected with series of events (with discussion). J Roy Stat Soc B 17:129–164
6. Cox DR, Lewis PW (1966) The statistical analysis of series of events. Methuen, London
7. Cox DR (1962) Renewal theory. Science Paperbacks and Methuen & Co. Ltd., London
8. Dorado C, Hollander M, Sethuraman J (1997) Nonparametric estimation for a general repair model. Ann Stat 25:1140–1160
9. Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall/CRC, Boca Raton
10. Jardine AKS, Tsang AHC (2006) Maintenance, replacement and reliability theory and application. CRC/Taylor & Francis, Boca Raton
11. Kijima M (1989) Some results for repairable systems with general repair. J Appl Probab 26:89–102
12. Lindqvist BH (2006) On the statistical modeling and analysis of repairable systems. Stat Sci 21(4):532–551
13. Pham H, Wang H (1996) Imperfect maintenance. Eur J Oper Res 94:425–438
14. Rigdon SE, Basu AP (2000) Statistical methods for the reliability of repairable systems. Wiley, New York
15. Team RDC (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
16. Yañez M, Joglar F, Modarres M (2002) Generalized renewal process for analysis of repairable systems with limited failure experience. Reliab Eng Syst Saf 77:167–180

A Toolkit Towards Performance Based Green Retrofit of HVAC Systems: Literature Review and Research Proposal

Shuo Chen, Guomin Zhang and Sujeeva Setunge

Abstract It has been argued that energy and environmental performance of existing Heating, Ventilation, and Air Conditioning (HVAC) systems can be improved significantly if the retrofit measures are selected and implemented properly. Various new retrofit technologies for HVAC systems have emerged that aims to reduce energy consumption and greenhouse emissions, but outcomes of retrofits often present difference from expected performance, with some succeeded, and some failed to meet the expected targets. This is, to some extent, due to relatively less knowledge and rare experience for design and construction of these technologies in HVAC systems. Moreover, lack of systematic assessment approach and limited examination guide make it hard to define the behaviour of each component of HVAC systems. These knowledge gaps and practical deficiencies have in the past prevented practitioners from selecting appropriate HVAC retrofit measures. Therefore, more in-depth research with practical case studies is needed to help test the retrofit outcomes and validate the claimed potential benefits. This chapter firstly will present an overview of the research on HVAC retrofits and introduce the challenges and risks commonly encountered in building renovation. Secondly, the key factors affecting HVAC retrofits will be reviewed and categorized. Finally, a brief description of the overall framework of the research method will be provided, which aims to develop a toolkit for green HVAC systems retrofits.

S. Chen (✉) · G. Zhang · S. Setunge
RMIT University, GPO Box 2476, Melbourne VIC 3001, Australia
e-mail: shuo.chen@rmit.edu.au

G. Zhang
e-mail: kevin.zhang@rmit.edu.au

S. Setunge
e-mail: sujeeva.setunge@rmit.edu.au

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_149

1 Introduction

In contemporary buildings, the heating, ventilation and air-conditioning (HVAC) system is an essential building service system, which provides a comfortable indoor environment for people to live and work [1]. Research on building energy usage found that HVAC systems alone generally account for 25–30 % of the total building energy usage [2]. Due to the increasing energy consumption, retrofitting of HVAC systems in the office building has gained more interest in the past few years [3, 4, 5]. Innovative work has been carried out by architectural and engineering groups to retrofit HVAC in office buildings and to lower carbon emission [6, 7, 4, 8]. Efforts have been made by scholars and industry practitioners in creating various new retrofitting technologies for HVAC systems, which are aiming at reducing energy cost and improving user comfort [9, 10, 11]. However, the outcomes of retrofitted HVAC systems present difference from expected performance, with some succeeded and some failed to meet the expected targets [12, [13]. This has given rise the so-called "gap" between the expected and actual performance of retrofitted HVAC, which has been widely documented in several studies [14, 15, 16]. In order to close this gap between expected and actual performance in a real HVAC retrofitting project, the property owners and asset managers commonly encounter uncertainties and risks which have been regarded as considerable technical challenge.

1.1 Challenges for HVAC Retrofits

Throughout reviewing the literature, the challenges mainly came from following aspects.

Insufficient real HVAC retrofit project has been implemented. There is a large body of research on HVAC retrofits available in the public domain. But, existing HVAC continue to be upgraded at a very low rate [12]. For instance, existing commercial HVAC stock is currently being retrofitted at a rate of approximately 2.2 % per year only [17]. Most previous studies were carried out using numerical simulations. Actual energy savings due to the implementation of retrofit measures in real buildings may be different from those estimated. More research with practical case studies is needed to help test the retrofit outcomes and validate the claimed potential benefits.

While many of these so-called green technologies have been invented in order to save operational energy, a review of HVAC retrofits literature indicates that few studies have been conducted to understand the advantages or disadvantages of green retrofits technologies [4, 18, 9]. As relatively less experience has been accumulated for design and construction of those new retrofit measures, it is hard to ascertain their suitability in different buildings. These knowledge gaps and practical

deficiencies have in the past prevented practitioners from selecting appropriate HVAC retrofit measures.

HVAC is sophisticated mechanical equipment and all components require retrofits through various forms of restoration. Lack of systematic assessment approach and limited examination guide information make it hard to define the behavior of each component of HVAC systems. Dedicating efforts to develop comprehensive pre and post retrofit assessments model will be demanded. Those will help property owners and asset managers access a comprehensive list of criteria to evaluate current HVAC systems and select performance-based retrofits.

Decision as to which retrofit technology (or measure) should be used for a particular project is a multi-objective optimization problem [14, 19, 20, 18, 21]. Insufficient attention has been paid to criteria such as human comfort, environmental sustainability and energy efficiency, which are not easily expressed or quantified. In many cases, advanced HVAC systems that prioritize cost savings are generally chosen, which may lead to a biased selection process [22, 23, 13]. The optimal solution is a trade-off among a range of energy related and non-energy related factors, such as energy, economic, technical, environmental, regulations, social, etc. [17, 24].

These knowledge gaps and practical deficiencies have in the past prevented practitioners from achieving performance-based HVAC retrofits. They have not been able to access a comprehensive list of criteria to evaluate the performance of HVAC, and there has also been a lack of a rational and systematic approach to facilitate the selection of performance-based retrofits. The lack of research into the process of performance-based retrofits selection and the resulting inefficiency of the selection evaluation approach would possibly lead to a less than optimal selection of HVAC retrofits measures, which might fail to satisfy the expectations of developers and clients.

1.2 Objectives of Study

With the limitations and deficiencies of the current research in mind, this study will aim to

- Capture the processes of retrofit and specific sustainable features considered, adopted and implemented in HVAC systems.
- Evaluate the pre and post retrofit of HVAC systems from economic, energy, users' satisfactions and environmental impact perspectives.
- Determine the effects of performance targets on retrofit solutions.
- Optimize the retrofit methods and make decisions on retrofit solutions to better meet performance targets.
- Develop a toolkit capturing outcomes above, which can be used to select sustainability features in a green HVAC retrofit for an expected level of performance.

2 A Review of Key Factors Affecting HVAC Retrofits

As introduction in the preceding section, there has been no systemic research to investigate the critical factors of HVAC retrofit in delivering green projects. However, there are many lists of critical factors for construction projects introduced by various researchers in the previous decades. Ma et al. [17] maintained that success of a building retrofit programme is determined by six aspects, namely, policies and regulations, client resources and expectations, retrofit technologies, building specific information, human factors and other uncertainty factors. Belassi and Tukel [25] classified the factors into five distinct groups according to which element they relate to the project manager, the project team, the project itself, the organization, and the external environment. The classification of alternative technical solutions concerning the HVAC's design are economic, energy and environmental criteria as well as criteria of user's satisfaction [19]. Chan et al. [26] identified five groups of factors, namely, project-related factors, procurement-related factors, project management factors, project participants-related factors, and external factors. All the above classification methods have some similarity. Our list of proposed factors was derived from an extensive literature review. In general, the critical factors could be divided into four categories: stakeholders' factors, sustainable factors, policy and regulation factors, and retrofit planning and technologies' factors.

2.1 Stakeholders' Factors

The key role of the client is to create an organisational climate that encourages green retrofits. To foster this, certain capabilities and attitudinal aspects of clients are necessary. The client's characteristics are shown in Table 1. To meet the demanding requirements of specific projects, contractors and technicians must have characteristics in Table 2.

2.2 Sustainable Factors

Cost effectiveness is regarded as a key factor in selecting the components for a retrofit project [27]. The cost factor is considered the main concern of building

Table 1 Stakeholders' factors—clients' characteristics

Clients' characteristics	References
Awareness of green HVAC retrofits outcome	Xu et al. [24]
Competence of HVAC technical knowledge	Mitchell [12]
Ability to contribute ideas to HVAC retrofits design process	Chan et al. [26]
Ability to contribute ideas to HVAC retrofit construction process	Chan et al. [26]
Skilled leadership of organizing HVAC retrofits project	Xu et al. [24]

Table 2 Stakeholders' factors—HVAC contractors and HVAC technicians' characteristics

HVAC contractors and HVAC technicians' characteristics	References
Commitment of personnel and equipment resources	Mitchell [12]
Commitment of on-time and on-budget project delivering, no variations	Mitchell [12]
Be versed in HVAC design related policies and regulations	Mitchell [12]
Be versed in new HVAC design technologies for green outcomes	Mitchell [12]
Willingness to understand technical drivers of green outcomes	Xu et al. [24]
Past-related experience of using new HVAC design technologies for green outcomes	Mitchell [12]
Diversity of technical capability, capability of personnel	Chan et al. [26]
Size of company's organizations and number of subcontractors engaged on project	Chan et al. [26]

developers as they want to search for ways to reduce costs of retrofitting HVAC, thus increasing its value. The factors included in the economic aspect show in Table 3.

In recent years, increasing anthropogenic carbon emissions have been recognised as a cause of global climate change. A number of studies have identified buildings as being responsible for about half of all energy consumption, and, in turn, as responsible for about half of the greenhouse effect due to carbon dioxide emissions [28]. This has aroused a growing awareness of the need for energy efficiency in the design of the modern buildings [29]. Of all the building services

Table 3 Sustainable factors—economic factors

Economic factors	References
First cost—including labor, materials and equipment costs	Buys and Mathews [27], Avgelis and Papadopoulos [19]
Low operating and maintenance cost	Baek and Park [3], Dascalaki and Santamouris [4], Harris et al. [5], Avgelis and Papadopoulos [19]
Low life cycle cost	Dascalaki and Santamouris [4], Harris et al. [5]
Reducing operating cost through investing in energy efficiency equipment	Buys and Mathews [27]
Reducing vacancy rates and improving HVAC design which allows for higher flexibility	Buys and Mathews [27], Dascalaki and Santamouris [4]
Short pay-back period	Alwaer and Clements-Croome [30], Avgelis and Papadopoulos [19]
Higher rental income	Asadi et al. [14], Avgelis and Papadopoulos [19], Chidiac et al. [20], Doukas et al. [18] and Guo et al. [21]
Higher overall capital value of the building	Chidiac et al. [20] and Doukas et al. [18]

Table 4 Sustainable factors—environmental factors

Environmental factors	References
Lower energy consumption and lower carbon emissions	Avgelis and Papadopoulos [19], Ellis and Mathews [29], Asadi et al. [14], Dascalaki and Santamouris [4], Harris et al. [5]
Reducing pollution related to fuel consumption	Avgelis and Papadopoulos [19], Ellis and Mathews [29], Asadi et al. [14], Dascalaki and Santamouris [4], Harris et al. [5]

Table 5 Sustainable factors—social factors

Social factors	References
Improving user comfort: thermal comfort, humidity, and noise level	Avgelis and Papadopoulos [19]
Improving indoor air quality	Avgelis and Papadopoulos [19]
Reducing contamination and odors emissions	Alwaer and Clements-Croome [30]
Health, absenteeism, and productivity	Dascalaki and Santamouris [4]
Future-proofing against tenant demands and government regulations	Dascalaki and Santamouris [4]
Improving corporate image	Dascalaki and Santamouris [4]
Making the building more attractive to investors	Guo et al. [21]
Making the building more attractive to high quality tenants	Guo et al. [21]
Benefits for the city	Guo et al. [21]

concerned, HVAC systems are regarded as the most energy-intensive. The factors included in the environmental aspect show in Table 4.

A complement to cost and environmental factors, the basic intention of a HVAC to be planned, designed, built and managed is to offer an environment in which occupants can carry out their work, feel well and to some extent feel refreshed [30]. A truly green HVAC must address occupant well-being and health, and needs to take the quality of the working and living environment into account when bringing in new technology for the purpose of improving the performance of business organisations. Thus, The factors included in the social aspect show in Table 5.

2.3 Retrofit Planning and Technologies' Factors

Retrofit technologies are energy conservation measures used to promote building energy efficiency and sustainability. Retrofit technologies range from the use of energy efficient equipment, advanced controls and renewable energy systems to the

Table 6 Retrofit planning and technologies' factors

Retrofit planning and technologies' factors	References
Selecting of appropriate procurement strategies	Mitchell [12]
Coordination of HVAC retrofits and the whole building retrofits	Mitchell [12]
Detailed understanding of the HVAC's current state	Ma et al. [17]
Completeness of documentation and communication	Ma et al. [17] and Mitchell [12]
Building specific information	Ma et al. [17]
Geographic location, Building type, *size*, *age*, *Service systems*	
Occupancy schedule, Operation schedule and maintenance records	
Availability of green HVAC retrofit technologies	Ma et al. [17] and Mitchell [12]
Comprehensibility of newer HVAC technologies	Ma et al. [17] and Mitchell [12]
Clarity of HVAC performance assessment and diagnostic	Ma et al. [17] and Mitchell [12]
Completeness of design retrofit programming and accuracy of retrofit design analysis	Ma et al. [17]
Features of selected HVAC systems	Wong and Li [2]
Long life span, *Ability of further upgrade*, *Flexibility of control*	
Compatibility with other building systems, *Integrated with building automation systems*	

changes of energy consumption patterns, and the application of advanced heating and cooling technologies. Retrofit measures should be considered in their order of economic payback, complexity and ease of implementation. The effectiveness of a building retrofit is also dependent on building-specific information, such as geographic location, building type, size, age, occupancy schedule, operation and maintenance, energy sources, utility rate structure, building fabric, services systems, etc. For a particular project, the optimal retrofit solutions should be determined by taking into account building specific information. Thus, the main factors of this aspect are shown in Table 6.

2.4 Policy and Regulation Factors

Policy and regulations are energy efficiency standards, which set minimum energy efficiency requirements for retrofitting of existing HVAC systems. Governments may provide financial support and subsidies to assist building owners and developers in achieving the required energy performance targets through implementing energy retrofit measures. Thus, the split incentives increase the willingness of

Table 7 Policy and regulations' factors

Policy and regulations' factors	References
Availability of renovation policies and political strategies promote housing renovation	Ma et al. [17]
Closing the gap between environmental quality in building stock and NABERS, Green Star or ABGR environmental ratings standards	Ma et al. [17]
Incentives form government	Ma et al. [17]

building owners to pay for retrofit, which offer great opportunities for improved energy efficiency, increased staff productivity, reduced maintenance costs and better thermal comfort. The main factors of this aspect are shown in Table 7.

3 Proposed Research Method

The overall research consists of four major phases: Firstly, develop a conceptual framework of performance based green HVAC retrofit. Based on the review of existing literature, a proposed systematic retrofit framework will be developed, which contains major factors influencing the retrofit failures. Secondly, to develop and test the conceptual model, a questionnaire survey, including a rating method will be undertaken. Surveys are considered as the most feasible and adequate research strategy in this study as it is appropriate to deal with the questions of 'what' the factors are, and 'how much' weight these factors have. The rating method uses an online questionnaire, sent to a large group of building experts and professionals who have the knowledge and experience of intelligent buildings, to collect data and identify a group of factors for the HVAC retrofit. Then, through the questionnaire sent to the group of experts, the Statistical Package for the Social Sciences (SPSS) method will be adopted to test the comparability of the factors. Their mean weights are computed to prioritize or rank the factors and distinguish the most important factors from the least important ones. The next stage of research involves the investigation of solutions. The semi-structured interviews will be conducted to further explore the industry practitioners' opinion on the factors of HVAC retrofit. The research will categorize these potential barriers into potential influence factors pertaining to HVAC planning, design, installation, and operation and maintenance. Then, interviewees will be invited to make comments in order to pinpoint the significant influence factors and investigate the appropriate solutions. The optimal solution will be a trade-off among a range of energy related and non-energy related factors, such as energy, economic, technical, environmental, regulations, and social. Finally, to validate the practical mode, a HVAC retrofitting project is introduced as a case study to validate the significant influence factors and appropriate solutions concluded by the questionnaire and interview surveys. It aims

to improve the retrofitting strategies at different stages of HVAC retrofit planning, design, installation, and operation and maintenance. Those findings will be integrated to improve the proposed HVAC retrofit framework.

4 Conclusions

This article provides a review on the existing HVAC retrofits literature, highlights the research gaps and points out new research directions. Previous studies on HVAC retrofits have contributed to our understanding of the current risk and challenges encountered in the HVAC construction industry. Little research has studied how to rationally and systematically select performance-based retrofit methods. In addition, the review also identifies that the resulting inefficiency of the selection evaluation approach would possibly lead to a less than optimal selection of HVAC retrofits measures, which might fail to satisfy the expectations of developers and clients. These limitations and challenges in HVAC retrofitting research and practice provide rich research opportunities in this area.

Following comprehensive literature review, 46 nominated factors affecting HVAC retrofits were established. They were categorized into four main aspects, namely stakeholders' factors, sustainable factors, policy and regulation factors, and retrofit planning and technologies' factors. For next stage of this Ph.D. study, a research survey will be conducted to analyze the significance of these above factors and investigate appropriate retrofit solutions. The results collected through research survey will contribute to developing dedicated HVAC retrofit assessments model and establishing a toolkit of green retrofits design. The findings of the research will provide valuable information to asset managers and help them undertake performance-based retrofits.

References

1. So ATP, Chan WL (1999) Interlligent Building Systems. Kluwer Academic Publishers, Boston
2. Wong JKW, Li H (2010) Construction, application and validation of selection evaluation model (SEM) for intelligent HVAC control system. Autom Constr 19:261–269
3. Baek C-H, Park S-H (2012) Changes in renovation policies in the era of sustainability. Energy Build 47:485–496
4. Dascalaki E, Santamouris M (2002) On the potential of retrofitting scenarios for offices. Build Environ 37:557–567
5. Harris J, Anderson J, Shafron W (2000) Investment in energy efficiency: a survey of Australian firms. Energy Policy 28:867–876
6. Ascione F, de Rossi F, Vanoli GP (2011) Energy retrofit of historical buildings: theoretical and experimental investigations for the modelling of reliable performance scenarios. Energy Build 43:1925–1936

7. Chidiac SE, Catania EJC, Morofsky E, Foo S (2011) A screening methodology for implementing cost effective energy retrofit measures in Canadian office buildings. Energy Build 43:614–620
8. Gamtessa SF (2013) An explanation of residential energy-efficiency retrofit behavior in Canada. Energy Build 57:155–164
9. Juan Y-K, Gao P, Wang J (2010) A hybrid decision support system for sustainable office building renovation and energy performance improvement. Energy Build 42:290–297
10. Ma Z, Wang S (2011) Online fault detection and robust control of condenser cooling water systems in building central chiller plants. Energy Build 43:153–165
11. Rysanek AM, Choudhary R (2012) A decoupled whole-building simulation engine for rapid exhaustive search of low-carbon and low-energy building refurbishment options. Build Environ 50:21–33
12. Mitchell M (2009) HVAC retrofitting for green refurbishments in occupied buildings, p. 17
13. Zmeureanu R (1990) Assessment of the energy savings due to the building retrofit. Build Environ 25:95–103
14. Asadi E, da Silva MG, Antunes CH, Dias L (2012) Multi-objective optimization for building retrofit strategies: A model and an application. Energy Build 44:81–87
15. Azar E, Menassa CC (2012) A comprehensive analysis of the impact of occupancy parameters in energy simulation of office buildings. Energy Build 55:841–853
16. Xing Y, Hewitt N, Griffiths P (2011) Zero carbon buildings refurbishment—a hierarchical pathway. Renew Sustain Energy Rev 15:3229–3236
17. Ma Z, Cooper P, Daly D, Ledo L (2012) Existing building retrofits: methodology and state-of-the-art. Energy Build 55:889–902
18. Doukas H, Nychtis C, Psarras J (2009) Assessing energy-saving measures in buildings through an intelligent decision support model. Build Environ 44:290–298
19. Avgelis A, Papadopoulos AM (2009) Application of multicriteria analysis in designing HVAC systems. Energy Build 41:774–780
20. Chidiac SE, Catania EJC, Morofsky E, Foo S (2011) Effectiveness of single and multiple energy retrofit measures on the energy consumption of office buildings. Energy 36:5037–5052
21. Guo B, Belcher C, Roddis WMK (1993) RetroLite: An artificial intelligence tool for lighting energy-efficiency upgrade. Energy Build 20:115–120
22. Mahlia TMI, Razak HA, Nursahida MA (2011) Life cycle cost analysis and payback period of lighting retrofit at the University of Malaya. Renew Sustain Energy Rev 15:1125–1132
23. Rey E (2004) Office building retrofitting strategies: multicriteria approach of an architectural and technical issue. Energy Build 36:367–372
24. Xu P, Chan EH-W, Qian QK (2011) Success factors of energy performance contracting (EPC) for sustainable building energy efficiency retrofit (BEER) of hotel buildings in China. Energy Policy 39:7389–7398
25. Belassi W, Tukel OI (1996) A new framework for determining critical success/failure factors in projects. Int J Project Manage 14:141–151
26. Chan A, Scott D, Chan A (2004) Factors affecting the success of a construction project. J Constr Eng Manage 130(1):153–155
27. Buys JH, Mathews EH (2005) Investigation into capital costs of HVAC systems. Build Environ 40:1153–1163
28. Li J, Colombier M (2009) Managing carbon emissions in China through building energy efficiency. J Environ Manage 90:2436–2447
29. Ellis MW, Mathews EH (2002) Needs and trends in building and HVAC system design tools. Build Environ 37:461–470
30. Alwaer H, Clements-Croome DJ (2010) Key performance indicators (KPIs) and priority setting in using the multi-attribute approach for assessing sustainable intelligent buildings. Build Environ 45:799–807

Risk Management Based on Probabilistic ATC Under Uncertainty

Mengqi Li and Minghong Han

Abstract This chapter presents a methodology for analytical target cascading (ATC) under uncertainty to address the risk management problem. The proposed hierarchical ATC structure is exactly corresponding to the systematic risk management, which is a multidisciplinary optimization procedure. Since the uncertainty induces risks, the proposed probabilistic algorithm reformulates the ATC method by setting random variables and probabilistic constraints. The proposed ATC method decomposes risk management problem into hierarchical sub-problems, which are linked directly above and below using mean values and standard deviations. With the given risk targets from upper levels transmitting downward, each sub-problem at each level of the hierarchy operates the adaptive optimization method to narrow the gaps between responses and the distributed targets. Once the convergence is attained by iterating between top and bottom, variables and parameters are optimized to reduce the risks. The Risk can be regarded as an optimization target together with efficiency and cost, or it can be contained in constraints in each sub-problem to optimize the efficiency and cost within the prescribed risk boundary. A case of risk management optimization is given to verify the proposed methodology. The results confirm the applicability and efficiency of the probabilistic ATC method under uncertainty in risk management.

1 Introduction

Risk must be planned with the formation of decision-making, which is the deviation between consequence and anticipation from the decision maker. To avert or reduce the risk, a series of procedures, methods, technologies, implements and specifica-

M. Li · M. Han (✉)
School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China
e-mail: hanminghong@buaa.edu.cn

M. Li
e-mail: Lemonq@outlook.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_150

tions are established by definition, identification, analysis and assessment of risk management. Risk management is an effective measure to reduce costs and loss [1]. Systems are the most objects of risk management, and systematic risk is generated from the restriction of technology, cost, production, environment etc. in systematic design, procurement and deployment. In general, the systematic complexity increases the risk in systems [2].

Multidisciplinary design optimization (MDO) is an applied methodology to dispose the systematic risk management. MDO is a growing field of research with a wide range of applications. When optimizing engineering systems that involve multiple disciplines or systems, sequential optimization is often not able to find the true optimum of the systems. Thus it is important that interactions be properly accounted for, both in the solution of the coupled governing equations and the optimization. Only by considering these interactions during the optimization process can the true optimum of the coupled systems be determined [3]. For a large engineering systems, uncertainty is one of the dominating factors to induce risk [4]. People often used the empirical method or redundancy to cope with such problem in the past, however the result is unsatisfactory. To address the uncertainty in systems, reliability-based multidisciplinary optimization (RBMDO) is adopted, which could accurately arrange each design factor to achieve the holistic optimum [5]. Analytical target cascading (ATC) is one of the mature optimization structures in MDO, and it has been successfully applied in uncertainty-based optimization problem in some researches [6–8]. ATC is analogous to the risk management model in systems, since they are both constructed level by level, and the interaction between levels could be emulated. After modelling the risk management by ATC, some RBMDO methods are used in this chapter to give an optimal scheme of risk management.

2 Risk Management Systems and ATC

This chapter proposes the model of risk management to be built on the ATC. In this section, the risk management and ATC are reviewed, and we analyse the reason why they are suitable to be integrated together.

2.1 Risk Management Systems

Most objects of risk management are large complex systems, which constitute many interactive subsystems. Each subsystem has own respective target, and they coordinate to achieve the target in the uppermost level. Only by satisfying targets in each level, the risk of holistic systems can be reduced. Therefore risk management must consider layered structure, since large complex systems are hierarchical [9]. After superincumbent decomposition of demand from the highest level, risk assessments aggregate from the lower system to the top.

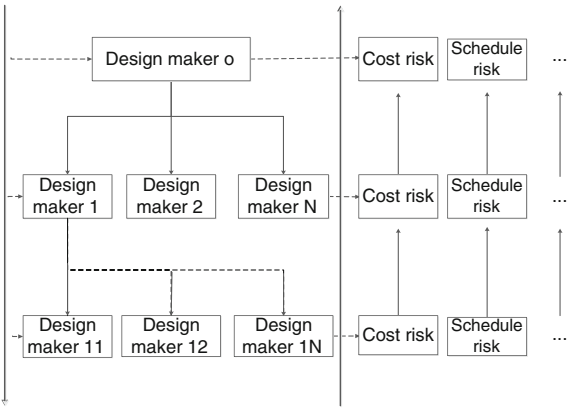


Fig. 1 Hierarchical risk management

The participants of risk management in systems involve acquisitions and undertakers. When operating systems, undertakers in subsystems should consider the demand from superiors such as system function, capability and assigned tasks, and they also must take development cost, progress and risk of their own into account. Risk, efficiency and cost should be placed on equal terms. Risk management considering hierarchy is shown in Fig. 1.

2.2 Analytical Target Cascading

ATC is a multilevel optimization method to be applied to MDO problems of large scale hierarchical systems, it is more applied to concurrent computational technology and its structure is flexible and extensional. The convergence of ATC has been proved, and the convergence rate is fast. ATC enables sub problems in each element of the hierarchy link to the sub problems directly above and below, which would converge by iterating repeatedly between top and bottom. It does not require extensive links among all the subsystems, so that the data relationship can be simplified. Another important reason using ATC is that the uncertainty propagation is ore clear and orderly in this hierarchy ATC construction [5].

ATC decomposes MDO problem into hierarchical sub-problems, which are linked directly above and below. In general, three levels ATC contains supersystem (Sup), system (S) and subsystem (SS). Each element in each level has two types of models: optimal design models P and analysis models r. Optimal design models use analysis models to evaluate supersystem, system and subsystem responses. Thus, analysis models take design variables, parameters and lower level responses as inputs and return the responses to upper-level design problems as output. A response is an output from an analysis model, and a linking variable is a common design variable between two or more design problems sharing the same parent design problem.

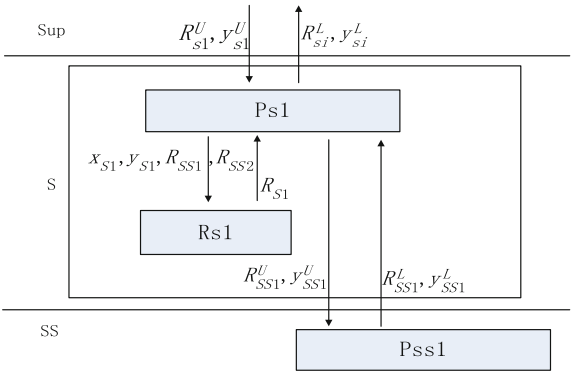


Fig. 2 Data flows from and into the system level element in ATC

For system S Targets for system responses $\mathbf{R}^U_{s,j}$ and system linking variables $\mathbf{y}^U_{s,j}$ are passed down from the supersystem level. After solving the system design problem, target values for system responses $\mathbf{R}^L_{s,j}$ and system linking variables $\mathbf{y}^L_{s,j}$ are passed up to the supersystem level. In this optimization process, the gap between variables and targets continually reduces and reaches convergence after limited cycles. $\varepsilon_{\mathbf{R}}, \varepsilon_{\mathbf{y}}$ are added to modify the deviation between target and response to coordinate the responses and linking variables from a lower level. At convergence, deviation tolerances become zero as the linking variables converge to the same values for the different design problems in the same level. For system S, responses from subsystem SS1 $\mathbf{R}_{ss1}$, system local design variables $\tilde{\mathbf{x}}_{s1}$, and system linking variables $\mathbf{y}_{s1}$ are input to the analysis model r_{s1}, whereas system responses Rs1 are returned as output. The data flow in element 1 in system level is shown in Fig. 2.

The problem at *j*th element of system level is stated as:

$$\begin{cases} \min & P_{s,j}: \ \|\mathbf{R}_{s,j} - \mathbf{R}^U_{s,j}\| + \|\mathbf{y}_{s,j} - \mathbf{y}^U_{s,j}\| + \varepsilon_{\mathbf{R}} + \varepsilon_{\mathbf{y}} \\ \text{w.r.t.} & \tilde{\mathbf{x}}_{s,j}, \boldsymbol{y}_{s,j}, \boldsymbol{y}_{ss}, \mathbf{R}_{ss}, \varepsilon_{\mathbf{R}}, \varepsilon_{\mathbf{y}} \\ \text{where} & \mathbf{R}_{s,j} = \mathrm{r}_{s,j}\left(\mathbf{R}_{ss}, \tilde{\mathbf{x}}_{s,j}, \boldsymbol{y}_{s,j}\right) \\ & \sum \|\mathbf{R}_{ss,k} - \mathbf{R}^L_{ss,k}\| \le \varepsilon_{\mathbf{R}}, \sum \|\mathbf{y}_{ss,k} - \mathbf{y}^L_{ss,k}\| \varepsilon_{\mathbf{y}} \\ & \mathbf{g}_{s,j}\left(\mathbf{R}_{s,j}, \tilde{\mathbf{x}}_{s,j}, \boldsymbol{y}_{s,j}\right) \le 0, \boldsymbol{h}_{s,j}\left(\mathbf{R}_{s,j}, \tilde{\mathbf{x}}_{s,j}, \boldsymbol{y}_{s,j}\right) = 0 \\ & \tilde{\mathbf{x}}^{min}_{s,j} \le \tilde{\mathbf{x}}_{s,j} \le \tilde{\mathbf{x}}^{max}_{s,j}, \quad \boldsymbol{y}^{min}_{s,j} \le \boldsymbol{y}_{s,j} \le \boldsymbol{y}^{max}_{s,j} \end{cases} \tag{1}$$

To decrease the risk, hierarchical decomposition systems risk management is proposed to be established. Such a model based on the ATC is shown in Fig. 3. Each supersystem, system and subsystem have optimal model and analysis model, in which each design maker in each element can select appropriate design and optimization method to operate risk assessment and control with the satisfaction of the targets transmitted from the upper level. The risk could be served as the

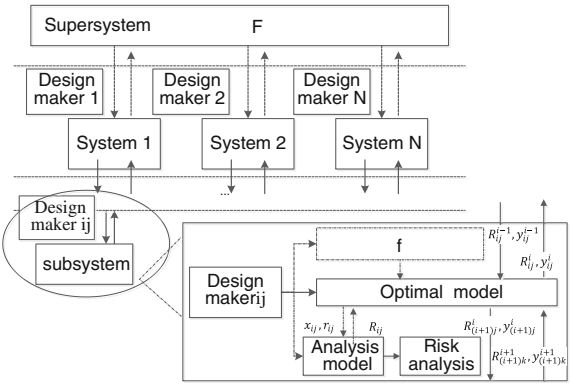


Fig. 3 Model of risk management based on ATC

constraint of cost and efficiency, which could be treated as the optimization of cost and efficiency within the prescribed risk boundary, and it can also be regarded as the optimization target.

3 Probabilistic ATC under Uncertainty

RBMDO can improve the system design by coordinating interactions, and meanwhile enhance the reliability to reduce the risk by taking uncertainties into account. Assuming that system design variables and parameters only have aleatory uncertainties, since other uncertainties such as epistemic uncertainty only exists when information is lacking. Based on the probability theory, the risk management problem is formulated as

$$\begin{cases} \max & F \\ \text{s.t.} & \mathrm{P_r}(f \le F) \le p_{f\_obj} \\ & \mathrm{P_r}(g \ge 0) \le p_{f\_con} \\ & \mathbf{x}^L \le u_{\mathbf{x}} \le \mathbf{x}^U \end{cases} \tag{2}$$

where $\boldsymbol{d}$ is the deterministic design variable vector. $\mathbf{x}$ is the random design variable vector. $\mathbf{x}^L$ and $\mathbf{x}^U$ are the lower and upper bounds of $\mathbf{x}$. The mean values of x are optimized to minimize the objective function f constrained by $g \le 0$. Due to the propagation of uncertainties, the responses of f and g are uncertain as well.

This method uses the first-order reliability method (FORM) to solve the Eq. (2), since the direct computation or Monte Carlo Simulations (MCS) of failure probability is inefficient. In FORM, probabilistic constraints can be expressed in two ways, which are reliability index approach (RIA) and performance measure approach (PMA). However, the convergence rate of RIA is slow, and errors are frequent when RIA operates [10]. Here we use more efficient PMA. The PMA formulation for probabilistic constraints is shown as:

$$\begin{cases} \min & g(\mathbf{u}) \\ s.t. & \|\mathbf{u}\| = \beta_t \end{cases} \tag{3}$$

$\mathbf{u}$ is the Gaussian random variable vector with zero mean and unit variance transmitted from $\mathbf{x}$ ($\mathbf{x} = \mathbf{u}_\mathbf{x} - \mathbf{u} \cdot \sigma_\mathbf{x}$). The optimum point on the given reliability surface is identified as the most probable point (MPP), β_t is the shortest distance from the origin to the point on the limit state surface in the standard normal space $\mathbf{U}$. According to the worst case analysis method, $\beta_t = 3$ when the reliability is 99.87 %.

To solve the PMA, advanced mean value (AMV) and conjugate mean value (CMV) are applied. When applied to a concave function, CMV is more efficient and stabilized whereas the AMV method tends to be slow in the rate of convergence or even divergent due to lack of updated information during iterative reliability analyses. However, AMV has fast convergence rate to solve the convex function. To develop an efficient MPP search method, the hybrid mean value is proposed, which combines AMV with CMV [11].

$$\begin{aligned} \mathbf{n}^{(k)} &= -\frac{\nabla_u g(\mathbf{u})^{(k)}}{\| \nabla_u g(\mathbf{u})^{(k)} \|} \\ \varsigma^{(k+1)} &= \left(\mathbf{n}^{(k+1)} - \mathbf{n}^{(k)}\right).\left(\mathbf{n}^{(k)} - \mathbf{n}^{(k-1)}\right) \end{aligned} \tag{4}$$

The type of performance function must be identified: if $\varsigma^{(k+1)} > 0$, the performance function at $\mathbf{u}_{\mathrm{HMV}}^{(k+1)}$ is convex, AMV should be used; else if $\varsigma^{(k+1)} \leq 0$, the performance function at $\mathbf{u}_{\mathrm{HMV}}^{(k+1)}$ is concave, CMV should be used. This method ensures the efficiency of searching MPP, and accurately describe the propagation of uncertainty.

After the $(r-1)$ cycle of optimization, the MPP of random variables is determined as $\mathbf{x}_{MPP}^{(r-1)^*}$ by PMA-HMV. At this point, some of the constraints may not be satisfied, then the MPP in this cycle are used to formulate the deterministic optimization problem for the next cycle, which would force the reliability to be improved by changing the values of $\boldsymbol{d}$. When every constraint and every deviation in the ATC are satisfied, convergence is achieved. Actually the distance between $\mathbf{x}$ and its MPP varies as the mean value of $\mathbf{x}$ changes. The MPP in the r cycle is formulated as

$$u_\mathbf{x}^{(r)} - \boldsymbol{s}^{(r)}, \boldsymbol{s}^{(r)} = u_\mathbf{x}^{(r-1)} - \mathbf{x}_{MPP}^{(r-1)^*} \tag{5}$$

Based on the equations above, the RBMDO problem of j element in system level in the r cycle is formulated as

$$\begin{cases} \max & F^{(r)}=\min \quad f_{s,j}\left(\boldsymbol{d}_{s,j}, u_{\mathbf{x}_{s\cdot j}}^{(r)}-\boldsymbol{s}_{s,j\_obj}^{(r)}\right) \\ \text{s.t.} & \mathrm{g}_{s,j,i}^{(r)}\left(\mathbf{d}_{s,j,i}, \mathrm{u}_{\mathbf{x}_{s\cdot j}}^{(r)}-\mathbf{s}_{s,j,i\_con}^{(r)}\right) \le 0 \\ & \boldsymbol{s}_{s,j\_obj}^{(r)}=u_{\mathbf{x}_{s\cdot j}}^{(r-1)}-\mathbf{x}_{MPP\_obj}^{(r-1)^*} \\ & \mathbf{s}_{s,j,i\_con}^{(r)}=u_{\mathbf{x}_{s\cdot j}}^{(r-1)}-\mathbf{x}_{MPP\_con}^{(r-1)^*} \end{cases} \tag{6}$$

The corresponding optimization problem in probabilistic ATC of typical three-level structure is formulated as

$$\begin{cases} \min & \|\mathbf{r}_{s,j}\left(\mathbf{d}_{s,j}, \mathbf{u}_{\mathbf{x}_{s,j}}^{(r)}, \mathbf{u}_{\mathbf{y}_{s,j}}^{(r)}, \mathbf{u}_{\mathbf{R}_{ss}}^{(r)}\right)-\mathbf{R}_{s,j}^{U}\|+\|\mathbf{u}_{\mathbf{y}_{s,j}}^{(r)}-\mathbf{u}_{\mathbf{y}_{s,j}}^{U}\|+\|\mathbf{d}_{\mathbf{y}_{s,j}}^{(r)}-\mathbf{d}_{\mathbf{y}_{s,j}}^{U}+ \\ & \varepsilon_{\mathbf{R}}+\varepsilon_{\mathbf{y}} \\ \text{s.t.} & \sum\left(\|\mathbf{d}_{\mathbf{R}_{ss,k}}-\mathbf{d}_{\mathbf{R}_{ss,k}}^{L}+\mathbf{u}_{\mathbf{R}_{ss,k}}^{(r)}-\mathbf{u}_{\mathbf{R}_{ss,k}}^{L}\|\right) \le \varepsilon_{\mathbf{R}} \\ & \sum\left(\|\mathbf{d}_{\mathbf{y}_{ss,k}}-\mathbf{d}_{\mathbf{y}_{ss,k}}^{L}+\mathbf{u}_{\mathbf{y}_{ss,k}}^{(r)}-\mathbf{u}_{\mathbf{y}_{ss,k}}^{L}\|\right) \le \varepsilon_{\mathbf{y}} \\ & \mathrm{g}_{s,j,i}^{(r)}\left(\mathbf{d}_{s,j,i}, \mathbf{u}_{\mathbf{x}_{s,j,i}}^{(r)}-\boldsymbol{s}_{\mathbf{x}_{s,j,i}\_\mathrm{con}}^{(r)}, \mathbf{u}_{\mathbf{y}_{s,j,i}}^{(r)}-\boldsymbol{s}_{\mathbf{y}_{s,j,i}\_\mathrm{con}}^{(r)}, \mathbf{u}_{\mathbf{R}_{ss,i}}^{(r)}-\boldsymbol{s}_{\mathbf{R}_{ss,i}\_\mathrm{con}}^{(r)}\right) \le 0 \\ & \mathbf{d}^{min} \le \mathbf{d} \le \mathbf{d}^{max}; \mathbf{x}^{min} \le \mathbf{u}_{\mathbf{x}}^{(r)} \le \mathbf{x}^{max}; \mathbf{y}^{min} \le \mathbf{u}_{\mathbf{y}}^{(r)} \le \mathbf{y}^{max}; 1 \le i \le Ng \end{cases} \tag{7}$$

$\boldsymbol{s}_{\mathbf{x}_{s,j,i}\_\mathrm{con}}^{(r)}$, $\boldsymbol{s}_{\mathbf{y}_{s,j,i}\_\mathrm{con}}^{(r)}$, $\boldsymbol{s}_{\mathbf{R}_{ss,\kappa}\_\mathrm{con}}^{(r)}$ are determined in the FORM process of constraints function with PMA-HMV method, and $\boldsymbol{s}_{\mathbf{x}_{s,j}\_\mathrm{con}}^{(r)}$, $\boldsymbol{s}_{\mathbf{y}_{s,j}\_\mathrm{con}}^{(r)}$, $\boldsymbol{s}_{\mathbf{R}_{ss}_\mathrm{con}}^{(r)}$hfill $\boldsymbol{s}_{\mathbf{R}_{ss,k}\_\mathrm{con}}^{(r)}$ are contained in $\boldsymbol{s}_{\mathbf{x}_{s,j,i}\_\mathrm{con}}^{(r)}$, $\boldsymbol{s}_{\mathbf{y}_{s,j,i}\_\mathrm{con}}^{(r)}$, $\boldsymbol{s}_{\mathbf{R}_{ss,\kappa}\_\mathrm{con}}^{(r)}$ respectively.

Differing from the probabilistic ATC formulation of decomposed subordinate elements, the optimization problem at supersystem is formulated as

$$\begin{cases} \min & \|\mathbf{r}_{\mathrm{sup}}\left(\mathbf{d}_{\mathrm{sup}}, \mathbf{u}_{\mathbf{x}_{\mathrm{sup}}}^{(r)}-\boldsymbol{s}_{\mathbf{x}_{\mathrm{sup}}\_obj}^{(r)}, \mathbf{u}_{\mathbf{R}_s}^{(r)}-\boldsymbol{s}_{\mathbf{R}_s\_obj}^{(r)}\right)-\mathbf{T}_{\mathrm{sup}}\|+\varepsilon_{\mathbf{R}}+\varepsilon_{\mathbf{y}} \\ \text{s.t.} & \sum\left(\|\mathbf{d}_{\mathbf{R}_{s,k}}-\mathbf{d}_{\mathbf{R}_{s,k}}^{L}+\mathbf{u}_{\mathbf{R}_{s,k}}^{(r)}-\mathbf{u}_{\mathbf{R}_{s,k}}^{L}\|\right) \le \varepsilon_{\mathbf{R}} \\ & \sum\left(\|\mathbf{d}_{\mathbf{y}_{s,k}}-\mathbf{d}_{\mathbf{y}_{s,k}}^{L}+\mathbf{u}_{\mathbf{y}_{s,k}}^{(r)}-\mathbf{u}_{\mathbf{y}_{s,k}}^{L}\|\right) \le \varepsilon_{\mathbf{y}} \\ & \mathrm{g}_{\mathrm{sup},i}^{(r)}\left(\mathbf{d}_{\mathrm{sup},i}, \mathrm{u}_{\mathbf{x}_{\mathrm{sup},i}}^{(r)}-\boldsymbol{s}_{\mathbf{x}_{\mathrm{sup},i}\_\mathrm{con}}^{(r)}, \mathrm{u}_{\mathbf{R}_{s,i}}^{(r)}-\boldsymbol{s}_{\mathbf{R}_{s,i}\_\mathrm{con}}^{(r)}\right) \le 0 \\ & \mathbf{d}^{min} \le \mathbf{d} \le \mathbf{d}^{max}; \mathbf{x}^{min} \le \mathbf{u}_{\mathbf{x}}^{(r)} \le \mathbf{x}^{max}; \mathbf{y}^{min} \le \mathbf{u}_{\mathbf{y}}^{(r)} \le \mathbf{y}^{max}; 1 \le i \le Ng \end{cases} \tag{8}$$

If the convergence efficient of optimization process is low, the standard deviation of responses and linking variables could be replenished in objective function and deviation tolerance [12]. The standard deviation of responses is

$$\sqrt{\sum\left(\frac{\partial \mathbf{r}(\mathbf{u})}{\partial \mathbf{u}}\right)^2 \sigma_{\mathbf{u}}^2} \tag{9}$$

In the optimization of the first cycle, each MPP could be estimated as the mean value of random variables, or can be given by an experienced engineer. After the

first optimization, the new generated MPP iterates in the next optimization process until the convergence is achieved. This is a sequential method to combine optimization and uncertainty analysis in the MDO, which avoids the inefficient nested optimization and analysis structure.

4 Example

The MDO formulation tested below is a benchmark of ATC [13]. Assuming that the optimization target is the schedule, the risks of laggard schedule, unexpected efficiency and exceeding cost are set as 10 % ($\beta_t = 1.28$). The problem has 12 deterministic positive design variables, 2 random variables, 4 equality and 6 inequality constraints. x_8, x_{11} are random variables with 0.01 standard deviations.

$$\begin{cases} \min \quad f = x_1^2 + x_2^2 \\ \text{s.t.} \quad \text{g1} : \frac{x_3^{-2}+x_4^2}{x_5^2} \le 1, \text{g2} : \frac{x_5^2+x_6^{-2}}{x_7^2} \le 1, \text{g3} : \frac{x_8^2+x_9^2}{x_{11}^2} \le 1 \\ \quad \text{g4} : \frac{x_8^{-2}+x_{10}^{-2}}{x_{11}^2} \le 1, \text{g5} : \frac{x_{11}^2+x_{12}^{-2}}{x_{13}^2} \le 1, \text{g6} : \frac{x_{11}^2+x_{12}^2}{x_{14}^2} \le 1 \\ \quad h1 : x_1^2 = x_3^2 + x_4^{-2} + x_5^2, h2 : x_2^2 = x_5^2 + x_6^2 + x_7^2 \\ \quad h3 : x_3^2 = x_8^2 + x_9^{-2} + x_{10}^{-2} + x_{11}^2, h4 : x_6^2 = x_{11}^2 + x_{12}^2 + x_{13}^2 + x_{14}^2 \end{cases} \tag{10}$$

This systematic risk management problem is decomposed into two levels:
Supersystem level P_{sup}:

$$\begin{cases} \min \quad x_1^2 + x_2^2 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 \\ \text{s.t.} \quad \left(x_{11} - x_{11}^{R_1^L}\right)^2 + \left(x_{11} - x_{11}^{R_2^L}\right)^2 \le \varepsilon_1 \\ \quad \left(x_3 - x_3^{R_1^L}\right)^2 \le \varepsilon_2, \quad \left(x_6 - x_6^{R_2^L}\right)^2 \le \varepsilon_3 \\ \quad \text{g1}, \ \text{g2}, \ h1, \ h2 \end{cases} \tag{11}$$

$$\text{System level} \ \ P_{S1} \ : \ \begin{cases} \min \quad \left(x_3 - x_3^{R_{\text{sup}}^U}\right)^2 + \left(x_{11} - x_{11}^{R_{\text{sup}}^U}\right)^2 \\ \text{s.t.} \quad \text{g3}, \text{g4}, \text{h3} \end{cases} \tag{12}$$

$$\text{System level} \ P_{S2} \ : \ \begin{cases} \min \quad \left(x_6 - x_6^{R_{\text{sup}}^U}\right)^2 + \left(x_{11} - x_{11}^{R_{\text{sup}}^U}\right)^2 \\ \text{s.t.} \quad \text{g5}, \text{g6}, \text{h4} \end{cases} \tag{13}$$

According to the previous described methods, MPP is determined after each optimization, and the targets $x_3^{R_1^U}, x_6^{R_2^U}, x_{11}^{R_{\text{sup}}^U}$ are transmitted down from the top level, and $x_3^{R_{\text{sup}}^L}, x_6^{R_{\text{sup}}^L}, x_{11}^{R_1^L}, x_{11}^{R_2^L}$ return as the feedback from the two systems. After iterating until the constraints and the deviations are satisfied, each design maker could know how to adjust the plan to reduce the risk of laggard schedule, and meanwhile

Table 1 Risk management results

	x_4	x_5	x_7	$\mathbf{u}_{\mathbf{x}_8}$	x_9	x_{10}	$\mathbf{u}_{\mathbf{x}_{11}}$	x_{12}	x_{13}	x_{14}
ATC	0.76	0.87	0.95	0.97	0.87	0.80	1.30	0.84	1.75	1.54
PATC	0.73	0.86	0.94	0.95	1.00	0.90	1.25	0.84	1.75	1.54

maintain the efficiency and cost. Table 1 displays the comparison of results from the normal ATC without considering the uncertainty and the proposed probabilistic ATC, which manifests the reasonableness of proposed risk method.

5 Conclusion

Risk is ubiquitous, only by the normative management can it be decreased. On the basis of characteristics in risk management, this chapter finds the past approaches cannot match the risk problem, and therefore does some research in the structure of risk management to build a model based on the hierarchy and multiple design makers. The Major purpose of this chapter is to integrate risk management with multidisciplinary optimization. To address the problem in the risk management, probabilistic analytical target cascading under uncertainty is applied to simulate the structure and the process in the risk management, and it is solved in the framework of reliability-based multidisciplinary optimization.

References

1. Power Michael (2007) Organized uncertainty: designing a world of risk management. Oxford University Press, Oxford
2. Porterl J, Squair MJ, Singh A (2006) Risk & safety aspects of systems of systems. In: 44th AIAA aerospace sciences meeting and exhibit, 2006, vol 16, pp 11677–11691
3. Joaquim RRAM, Chritopher M (2009) An object-oriented framework for multidisciplinary design optimization. In: AIAA 2009–1906
4. Chapman C, Stephen W (1996) Project risk management: processes, techniques and insights. John Wiley Press, UK
5. Yao W, Chen X, Wei Y, van Tooren M, Gao J (2011) Review of uncertainty-based multidisciplinary design optimization methods for aerospace vehicles. Prog Aerosp Sci 47:450–479
6. Hubin L, Wei C, Kokkolaras M (2006) Probabilistic analytical target cascading: a moment matching formulation for multilevel optimization under uncertainty. J Mech Des 28:991
7. Kokkolaras M, Moulrlatos ZP, Papalambros PY (2006) Design optimization of hierarchically decomposed multilevel systems uncertainty. J Mech Des 125:124–130
8. Kuo-Wei L, Harrison MK, Christopher H (2006) Multilevel optimization considering variability in design variables of multidisciplinary system. In: AIAA 2006–7061
9. Haimes YY (2005) Risk modeling, assessment, and management. Wiley, New York

10. Ahn J, Kwon JH (2006) An efficient strategy for reliability-based multidisciplinary design optimization using BLISS. Struct Multidisc Optim 31:363–372
11. Youn BD, Choi KK (2003) Hybrid analysis method for reliability-based design optimization. ASME J Mech des 125(2):221–232
12. Xiong F, Yin X, Chen W, Yang S (2010) Engineering optimization. Enchanced probabilistic analytical target cascading with application to multi-scale design. Taylor Francis 42 (6):581–592
13. Kim HM, Michelena NF, Papalambros PY, Jiang T (2003) Target cascading in optimal system design. ASME J Mech Des 125:474–480

Systems Engineering Approach to Risk Assessment of Automated Mobile Work Machine Applications

Risto Tiusanen

Abstract Mobile work machines are today equipped with automatic functionalities. There are fully autonomous mobile work machines operating in factories, mines and harbor terminals. Needs for better productivity, better mobile work machine utility and higher work quality in work sites is driving the work process management towards automated production instead of improving the management of separate manual work machine operations. Safety risks related to automated mobile work machine applications are dependent on several operation condition factors such as mode of operation, level of automation, human–machine–interaction, performed work task and working environment. System safety approach that has been developed among others for aviation, military, process industry and railroad systems is introduced as a solution for safety risk management for complex mobile work machine applications. An application of systems engineering based approach, supporting tools and methods have been developed in close co-operation with mobile work machine manufactures, system suppliers and system end users for automated mobile work machine applications. This chapter describes the developed safety engineering methodology and some evaluation results based on experiences from industrial case studies.

1 Introduction

In several industrial sectors such as mining, construction, civil engineering, container handling and material handling, which are using mobile work machines, a new growing trend is to automate mobile work machine operations. In factories and warehouses automatic guided vehicles and similar automatic material handling machine systems have been used for years. In mining industry machine automation technology has the goals of speeding production, improving safety, and reducing costs. Mines and mineral processing plants develop integrated process control

R. Tiusanen (✉)
VTT Technical Research Centre of Finland, Tekniikankatu 1, 33101 Tampere, Finland
e-mail: risto.tiusanen@vtt.fi

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_151

systems capable to improve plant-wide efficiency and productivity. In container handling industry an on-going trend is the development of seaport container terminals to use automated container handling and transportation technology. In large container terminals manually driven cranes are going to be replaced by automated ones and automated guided vehicles are often used instead of manually operated work machines [1–4].

When the machines are remotely controlled and the machine control is developing towards machine fleet control and management, the focus on machine safety issues changes to system safety issues and the risk management issues of the whole worksite environment throughout the whole life cycle of the machinery system. In large-scale machine automation applications, safeguarding arrangements and safety-related functions are complicated and difficult to maintain. Such machinery applications can be compared with large process automation applications. Safety-related control functions in highly automated machine systems include multi-dimensional aspects such as the operator's actions, user interfaces, communication protocols and machine level control signals.

Aspects related to system safety have been issued among mobile machine manufacturers concerning automated machinery system development. Current safety regulations and standardization do not link machinery safety design tasks properly to the machinery life cycle phases and the machinery system R&D process phases. Traditional management of machine safety issues is said to be not enough. Systematic Top–Down approach for large machinery applications to support systems engineering is needed. The machinery safety design standard ISO 12100 [5] gives machine designers an overall framework and guidance to design machines that are safe for their intended use. In spite of these guidelines there is an increasing need for knowledge about how to specify system level safety and reliability requirements for unique complex mobile machine applications. There is also a need for new procedures on how to manage system safety and system reliability risks throughout the whole life cycle of the system. Today mobile machinery end users in many industrial sectors refer to functional safety requirements set in IEC 61508 [6] which are difficult to apply in complex machinery systems.

2 Systems Engineering Approach Aims to Manage the Whole System Entity

Systems engineering is said to have two main perspectives: the technical discipline that concentrates on the design and operation of the system, and the managerial discipline concentrating on the systems engineering and project management over the whole system life cycle. The principles of systems engineering can be applied to any system and it provide systems thinking that can be employed at all levels. Systems engineering tries to ensure that all essential aspects of a system are considered, and integrated into a whole [7].

Issues such as reliability, logistics, and coordination of different teams, requirements management, evaluation measurements, and other disciplines become difficult in large projects. Systems engineering covers among others distributed and networked work-processes, optimization methods, and risk management in such projects. Systems engineering links together technical and human-centred disciplines such as industrial engineering, automation engineering, job planning and project management. Use of modelling and simulation techniques is essential in systems engineering to create, demonstrate and validate assumptions or theories on a system and the interactions within the system and its subsystems [8]. Analysis methods that allow early detection of possible hazards, threats, human factors and technical failures, in system safety engineering, are integrated into the design process. Commonly used motivation for systems engineering approach is the practical knowledge that decisions made at the beginning of a project whose consequences are not clearly understood can have enormous implications later during system operation. Systems engineers should examine these issues and provide systematically reasoned information for the project management to able to make the critical decisions [7, 9].

2.1 Systems Engineering Process Guides Towards Top–Down Problem Solving

Systems engineering process is an essential part of systems engineering management activities in system development and design. Its objective is to provide a process that transforms requirements into system specifications, different levels of system architectures and system design baselines. Important added value for the system development is the engineering discipline that provides the control and traceability of the decisions made for the design and solutions so that they meet the original customer needs and specifications. The systems engineering process is an iterative and recursive top–down problem solving process. It has three main tasks: requirement analysis, functional analysis and allocation, and design synthesis connected with control, feedback and verification loops (Fig. 1).

The systems engineering process is meant to be implemented throughout all life cycle phases and especially in system development. System development proceeds through the development stages such as conceptual design level, producing the system concept description; system design level, producing the system requirement specification; and subsystem or component design level, producing the detailed characteristics and performance descriptions for the implementation of the subsystems and components [7, 10].

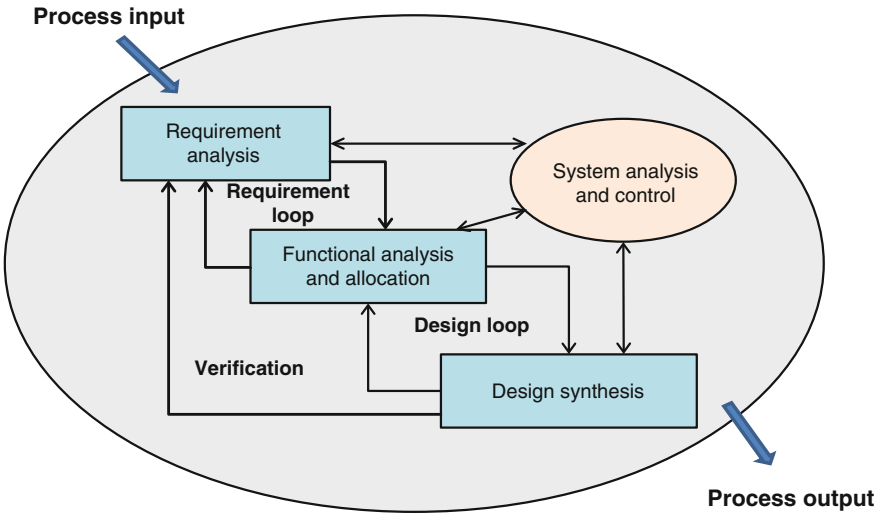


Fig. 1 Simplified flow chart of the systems engineering process according to [7]

2.2 *Integration of Safety Engineering into Systems Engineering Approach*

Risk management in systems engineering context deals with both risks related to the process of developing the system (project risks) and risks related to the system product (product risks). System level safety risks of a mobile work machine application under development belong to the latter category. The system safety concept means the effort to make things as safe as possible in the early stages of the system lifecycle by using engineering and management tools. It involves well planned, systematic safety analysis processes. Safety analysis means the recognition and improvement of dangerous features in a system. Hazards in a system should be identified and controlled before losses occur, with different analysis methods, at different stages at its life cycle [11, 12].

MIL-STD-882C [13] defines system safety and system safety engineering as follows: "The application of engineering and management principles, criteria, and techniques to optimize all aspects of safety within the constraints of operational effectiveness, time, and cost throughout all phases of the system life cycle." Safety is often taken into notice after an accident occurs and then corrections are performed to prevent similar accidents. The practical knowledge is that this approach is expensive and time consuming, also very inefficient, dangerous and often inhumane. The System Safety concept addresses the hazards before losses occur, making the system safe to operate and maintain. System safety engineering should be an essential part of the systems engineering process, so that the safety engineering aspects and safety requirement management do not separate from the main systems engineering

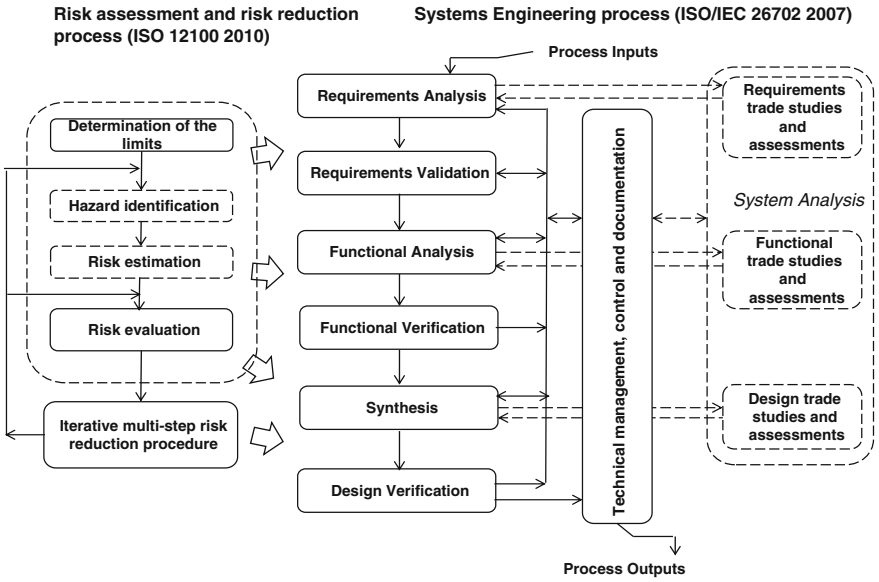


Fig. 2 Integration of risk assessment process from machinery sector and the general systems engineering processes

management agenda. Through analysis, design and management actions the hazards are identified, evaluated, eliminated and controlled in order to make the system safer (Fig. 2).

The basic machinery safety design standard ISO 12100 [5] gives machine designers an overall framework and guidance to design machines so that they are safe for their intended use. It specifies basic terminology, principles and a methodology for safety design of machinery. Procedures are described for identifying hazards and estimating and evaluating risks during relevant phases of the machine life cycle, and for the elimination of hazards or the provision of sufficient risk reduction. Guidance is also given for the documentation and verification of the risk assessment process.

3 System Safety Approach for Automated Mobile Machine Applications

New system level risk assessment approach for automated mobile machinery has been developed in VTT in co-operation with manufacturers, system suppliers and their subcontractors. The approach combines the base line hazard identification, task based hazard identification, system level HAZOP studies and risk estimation principles according to the system safety engineering concept (Fig. 3) [11–13]. The aim in hazard identification is to specify the automated system, its limits and

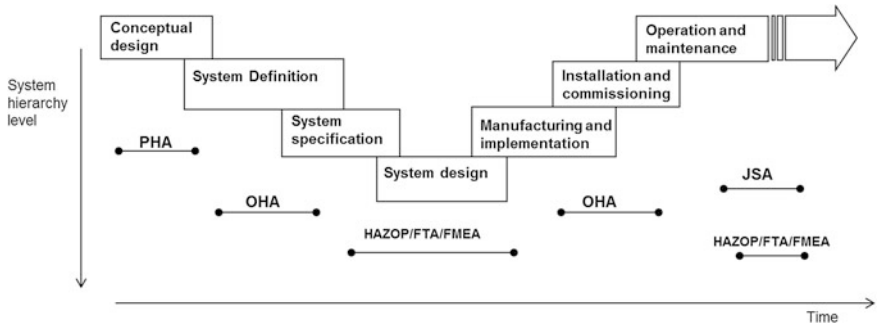


Fig. 3 Integration of system safety engineering activities in the system life cycle phases

interfaces, and identify all the potential hazards related to the automated machinery system in all its foreseeable operating situations. The intended use, related procedures and regular maintenance of the machine, as well as anticipated misuse of the machinery system should be taken into consideration in the identification of risk factors. Risk assessment and risk reduction is done following the procedure described in ISO 12100 [5].

The system safety engineering approach has been applied and evaluated in several automated mobile work machine applications such as:

- A semiautomatic ore loading and transportation system in an underground mine in Sweden.
- An autonomous ore transportation system in an underground mine in South Africa.
- An automatic crane system in a container terminal in Germany.
- An automated deposition machine to be used in a deep repository of waste nuclear fuel in Sweden.
- An automated harvester machine in Finland.

According to the practical experiences the outcome of Preliminary Hazard Analysis (PHA) and preliminary risk estimation form good baseline for the hazard list and preliminary safety requirements and conceptual solutions for overall safety measures in automated mobile machinery applications. The upper system level risk analysis of system operations and main system functionalities have been carried out using Operating Hazard Analysis (OHA) and two phase risk estimation procedure —before and after safety measures. Hazard and Operability study (HAZOP) methodology supported by function level drawings and database application for data collection and documentation seems to work out well for automation platform development purposes, customer specific application design verification purposes and for safety system validation purposes. The analysis of deviations and possible consequences in different system levels in HAZOP studies succeeded well in multi technological analysis teams.

The risk estimation methodology using risk matrixes and the guidelines for the interpretation of risk estimation results should be developed to support better the decision making in risk evaluation and trade studies in the various stages in systems engineering processes [7] (Fig. 2).

3.1 Simulator Assisted Safety Engineering

3-D modelling, system simulation and virtual environments are used commonly in systems engineering and machinery design. It is well known that simulator environments such as flight simulators, ship simulators, process and power plant simulators, and mobile machine training and R&D simulators offer a good possibility to develop, test and demonstrate different automation concepts with different system functionalities. HIL (Hardware-In-the-Loop) simulators have been developed and used successfully for mobile work machine control system and SW development for over 10 years. Current risk analysis and safety engineering practices in early life cycle phases of a machinery system typically use static system models such as 3-D models of machinery and facilities, preliminary layout drawings of the work site environment and functional descriptions. The use of machine simulators and virtual environments in risk analysis and safety engineering is still quite uncommon. According to [14] virtual environments combined with an analysis group improved the identification of critical safety situations in a machinery risk analysis. The use of virtual environments in plant design enhanced the development and analysis of different design variations from several points of view, including safety.

Research work on the simulator assisted safety engineering approach is going on in Finland in co-operation with mobile work machine manufacturers, machine control system designers, VTT, MTT Agrifood Research Finland and Technical University of Tampere. Objective of approach is to support system analysis and risk management in systems engineering process from the early system requirement specification through system design and verification up to the requirement validation phase. The research in this area has main goals: development of simulator and virtual reality environments to support hazard identification and risk estimation; and evaluation of safety concepts and safety functions using simulator and virtual reality environment (Fig. 4). The approach combines systems engineering approach, functional safety verification principles, safety engineering knowledge and modern simulator based design, engineering and verification environments. In practice the research and development effort integrates the virtual environment laboratory in VTT in Tampere, mobile machine simulator developed for this purpose in Creanex Oy in Tampere, and safety engineering process management tools and functional safety evaluations and calculation tools developed by VTT.

Experiences and results of the experiments conducting risk analysis with the simulator assisted methodology support the results of hazard identification and risk estimation in overall work site level and complex machinery applications described in [15]. In practice the methodology enables to visualize and study operational

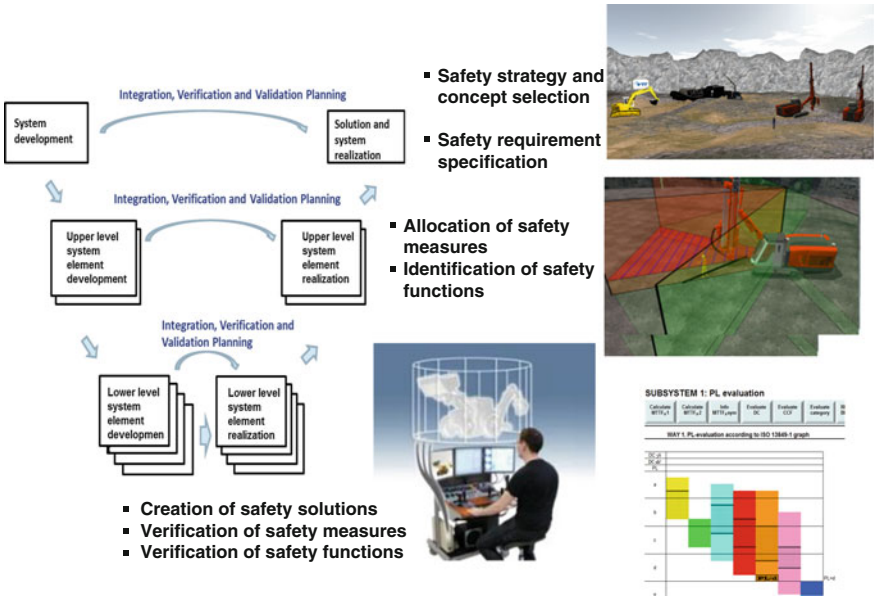


Fig. 4 Integration of simulator assisted safety engineering into the systems engineering process

situations that cannot be tested without damaging equipment or putting test persons in danger. The simulator assisted approach supports the systems engineering work and improves the current safety engineering practices for system analysis in conceptual design and system definition phases and also in the functional analysis and comparison of various concept and design alternatives [16].

3.2 Safety Engineering Data Management

Machinery safety design standard ISO 12100 [5] defines a process model for assessing risks of machinery, but it does not define the exact work flow how the risk assessments in a complex machinery automation and machine control system development process should be made. A model for risk assessment work flow and a safety engineering data model for risk assessment artifacts have been specified in co-operation with machine manufacturers and machine control system designers and suppliers. The objective has been a rigorous model to standardize the cooperation between the systems engineers and safety engineers. Also the deployment of third parties to carry out risk assessment tasks is supported [17]. In a long run the objective is that the database centric approach would help to harmonize the system safety engineering process and its interfaces between mobile work machine manufactures and their sub-contractors.

The proposed model has been tested by integration of an Application Life-cycle Management (ALM) tool with a database oriented risk assessment tool. The ALM tool is used to execute the workflows, to carry out version management and to provide the required traceability of artifacts [17]. The database tool has been used to carry out different types of risk analyses such as PHA, OHA or HAZOP studies. One big challenge in this kind of model is in the implementation of the tools that is used to utilize the model. The tools should be designed from the user's needs and make the risk assessment easier than before. The essential benefit and added value from the safety engineering data model is that the after the project in question is finished all the necessary safety related information such as the safety case or the technical file, is available for safety authorities with no or minimal add-on work.

3.3 Control System Functional Modelling and Simulation

To support the upper and lower system level risk analysis a new function level drawing technique has been developed in VTT. So called resource allocation drawing technique integrates information to a database from design documents like system (architectural) description, list of control and safety functions, design intent descriptions, module and I/O-lists and communication message specifications. These drawings depict the resources (sensors, actuators controllers, I/O-signals and communication signals) needed to perform the control function (Fig. 5).

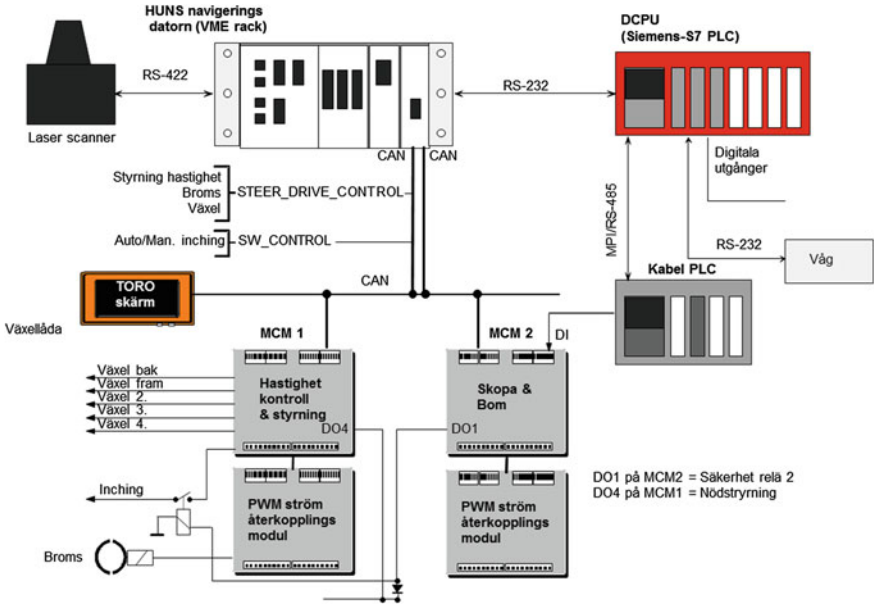


Fig. 5 An example of function level drawing of a distributed machine control system

Functional safety of machine control systems is related to the capability of a control system to implement the safety functions correctly. There are two machine control system safety design standards applicable in machinery sector concerning this topic. IEC 62061 [18] deals with electrical and programmable electronic machine control systems and defines safety integrity levels (SIL). ISO 13849-1 [19] defines safety performance levels (PL) and provides basis for the design and performance of safety related parts of machine control systems.

VTT has developed a PL calculation tool based on an Excel application according to the principles presented in ISO 13849-1 [19]. That makes it possible for designers to test the impact of different solutions and make safety related technical solutions based on this. The big challenge in such quantitative is how to get reliable and applicable source data for the calculation of functional safety parameters. At the present there are failure rates and other parameters available for very few components and safety devices [20]. VTT's PL calculation tool is semi-automatic. The values can be transferred from one design phase to other using macros. An Excel based tool has been developed that can also be used for documentation purposes. A safety block diagram forms the basis for PL calculations. Safety block diagram is not the same thing as reliability block diagram or functional diagram. A failure of some components has an effect on reliability but not on safety. In a safety block diagram all essential parts such as input devices, logics, output devices, which have an effect on the implementation of a safety function under study are presented. Typically, the safety block diagrams must be simplified in order to fit the structures into designated architectures of ISO 13849-1 [19, 20].

4 Discussion

The results of the safety engineering research and development work in the context of automated mobile work machine applications demonstrate that the integration of risk assessment approach into the systems engineering process provide applicable and systematically reasoned information for risk conscious decision making. The top–down proceeding safety engineering approach supports the sharing of system safety information and improves the common understanding of the system operations, human-system integration and interactions between subsystems.

System thinking clarifies that safety of complex mobile machine application cannot be solved by machine level safety solutions. A new way of combining system level and machine level analyses raises safety related issues for the planning and management of work, for co-operation inside the company, and for co-operation with other stakeholders such as subcontractors. System thinking supports also the specification of the risk reduction measures to all system levels considering all available types of safety measures: technical means, operational, managerial and organizational. As the implementation of functional safety procedure for safety related parts of the automation system IEC 61508 [6] seems to be de facto standard

also in mobile machinery sector. The system safety engineering approach and analysis methods and tools should be developed to support this common practice.

The use a simulator environment and visualization improved the conceptualization of the new work site environment, machine system operation and system functionalities. Simulator and virtual reality environments offer good possibilities to develop test and demonstrate different machinery automation concepts with different machine functionalities and in different work site situations. They also helped sharing of system information and improved the common understanding of the overall system and interactions between subsystems among the analysis team that consists of expert from different technological background. The simulator assisted approach seems to be able to provide new verifiable and traceable evidence for the possible certification of adaptive safety functionalities and safety systems in automated mobile work machine applications. The results in this context are in line with the general systems engineering principles and objectives for modelling and simulation technologies described in [7–10]. Modelling and simulation seems to be promising tool for the conceptual safety engineering. It provides possibilities to simulate and demonstrate various concepts and compare their effects on productivity and risk mitigation. The research in this field continues on aspects related to the evaluation criterion. Further studies focus on questions how to compare safety concepts and how to specify dimensions (parameters) for risk estimation and risk evaluation in human-machine interaction in complex mobile work machine application.

The risk assessment work flow model and a safety engineering data model for risk assessment artifacts support the systematic system safety engineering approach and a structured storage for system safety requirements, list of identified hazards, assessed risks and the specified risk reduction measures. Traceability from risk assessments to safety requirements—to design artifacts and—further to validation is documented and easy to maintain. The rigorous model supports the efforts to standardize the communication and cooperation interface between systems engineers and safety engineers. Also the deployment of third parties to carry out risk assessment tasks can be better coordinated.

The PL and SIL calculations of control system safety related functions only by hand are laborious and in practice computer based tools are necessary. The PL calculation tool developed for that purpose makes it possible for designers to analyze the impact of different control system safety principles and design and optimize the safety related functionalities and safety functions based on the standard based calculations. The biggest problems in the calculation of MTTFd values seem to be in the source data.

Acknowledgments The research work on system safety engineering approach and supporting methods and tools is part of on-going project FAMOUS (Future Semi-Autonomous Machines for Safe and Efficient Worksites). The research project is part of FIMECC's (Finnish Metals and Engineering Competence Cluster) research program EFFIMA (Energy and Life Cycle Efficient Machines) and the main financier of the research project is TEKES.

References

1. Ericsson M (2012) Mining technology—trends and development. POLINARES Consortium
2. Bellamy D, Pravica L (2011) Assessing the impact of driverless haul trucks in Australian surface mining. Resour Policy 36(2):149–338
3. Jämsä-Jounela S-L, Baiden G (2009) Automation and robotics in mining and mineral processing. Springer handbook of automation. Springer, Berlin, pp 1001–1013
4. Scott J (2012) Trends in marine terminal automation. Int J Port Technol 54:82–85
5. ISO 12100 (2010) Safety of machinery. General principles for design. Risk assessment and risk reduction
6. IEC-61508-1 (2010) Functional safety of electrical/electronic/programmable electronic safety-related systems—Part 1: general requirements. IEC
7. Defence Acquisition University (2001) Systems engineering fundamentals. Defence Acquisition University Press, Fort Belvoir, Virginia. http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-885j-aircraft-systems-engineering-fall-2005/readings/sefguide_01_01.pdf. Accessed 19 June 2013
8. Cellier FE, Floros X, Kofman E (2013) The complexity crisis—using modelling and simulation for system level analysis and design. In: The proceedings of the 60th anniversary seminar "Automation and systems without borders—beyond future" 21 May 2013 in Helsinki, Finland. The finnish society of automation. Helsinki 2013
9. INCOSE SE (2011) Handbook. Systems engineering handbook a guide for system life cycle processes and activities. INCOSE, San Diego
10. ISO IEC 26702 (2007) Systems engineering—application and management of the systems engineering process (IEEE STD 1220-2005); 1st edn
11. Ronald H, Moriarty B (1983) System safety engineering and management. Wiley, New York
12. Stephans R (2004) System safety for the 21st century. Wiley, New Jersey
13. MIL STD 882D (2000) Standard practice for system safety. Department of Defence
14. Määttä T (2003) Virtual environments in machinery safety analysis, vol 516. VTT Publications, Espoo, p 170
15. Tiusanen R, Malm T, Viitaniemi J (2013) Simulator assisted design approach for adaptive safety concepts in automated mobile work machine systems. In: The proceedings of the automation XX seminar "Automation and systems without borders—beyond future" 22 May 2013 in Helsinki, Finland. The Finnish Society of Automation. Helsinki 2013
16. Tiusanen R, Malm T, Ronkainen A (2012) Adaptive safety concepts for automated mobile work machine systems: simulator assisted research approach. In: Proceedings of the 7th international conference on the safety of industrial automated systems (SIAS 2012). IRSST. Montreal, 11–12 Oct 2012. http://www.irsst.qc.ca/en/sias2012.html
17. Alanen J, Tiusanen R, Sierla S, Papakonstantinou N, Koskinen K (2010) Rigorous work flow and data model for risk assessments of machine control systems. In: Proceedings of the 6th international conference on safety of industrial automation systems (SIAS 2010). Tampere, 14–15 June 2010. Finnish Automation Support Ltd
18. IEC 62061 (2005) Safety of machinery—functional safety of safety-related electrical, electronic and programmable electronic control systems. IEC
19. ISO 13849-1 (2006) Safety of machinery. Safety related parts of control systems. Part 1: general principles for design. ISO
20. Hietikko M, Alanen J, Malm T (2010) A safety process reference model and tool for the development of machine control systems. In: Proceedings of the 6th international conference on safety of industrial automation systems (SIAS 2010). Tampere, 14–15 June 2010. Finnish Automation Support Ltd

Stage Division Method and the Main Tasks of Products' Lifetime Cycle

Lei Gao, Ying Chen and Rui Kang

Abstract Due to the high requirement of the products' quality and updating speed, system engineering attracts broad attention. Great emphasis has been laid on the lifetime cycle process of system engineering and the theory of product's lifetime cycle. The products' lifetime cycle design, lifetime cycle management as well as the analysis and control of the lifetime cycle cost (LCC) all embody the theory of product's lifetime cycle. With this theory, we can not only get whole acquaintance with the product from the beginning, but also take into account the restraint and influence of the subsequent work, so that design flaws could be noticed in advance, time and cost could be reduced, quality could be improved and the competitiveness could be greatly enhanced. Stage division is an indispensable part of lifetime cycle theory. This chapter analyzes overseas and domestic research status as well as stage division method with its main tasks. Furthermore, it takes several concrete products as examples for introduction, including the weaponry, aircraft, automobile, vessel, construction and some others. At last, this chapter gives a simple analysis of inter-relationship between these stages and elaborates the theory of system engineering method from time dimension.

1 Introduction

With the rapid development of global economy and the sharp rise of science and technology level, modern manufacturing no longer produces single products as before. They tend to be numerous in varieties, complicated in structure and less time-consuming in making improvement. All these above mean the desire for small quantities and multiple species of the products. Engineering system theory emerges at this very moment and put forward the theory of products' lifetime cycle in time dimension, with which a great many practical engineering problems including time, costs, quality are well solved.

L. Gao (✉) · Y. Chen · R. Kang
Beihang University, No. 37, XueYuan Road, Haidian District, Beijing 100191, China
e-mail: gaoleijilin@126.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_152

As the market for products is highly competitive, developed countries realize very early that it is of significant importance to make effective use of existing enterprise resources. In 1968, they proposed to produce products which would only need one shot to conform to specifications, in order to fully utilize the resources and gain profits to the maximum. In 1969, American system engineering scientist Hall put forward the Hall three dimensions structure. It divides engineering technical program into time dimension, knowledge dimension and logic dimension [1]. The time dimension separates products' lifetime cycle into several stages, which is widely accepted. In 1986, the United States defense advisory committee reported to the President: "development cycle of the weapon system is too long, the production cost is too high but the performance is not quite satisfied." The U.S. Department of Defense instructed the Institute of Defense Analyze (IDA) to study the lifetime cycle design and the feasibility of applying it to weapons systems. The U.S. military production tends to use the theory of whole life cycle design, and also the test items of the weapon system, full scale development project and production projects. At the same time, many tools software supporting whole life cycle design are in research and development as well.

In our country, by organizing the implementation of the ninth five-year plan and the national 863 project, systems engineering and the whole life cycle design theory have had a good foundation. So many colleges and institutes dedicate to the research and have yielded fruitful results, which gain the favor of enterprises and have been widely used in the practice. Many related standards have been issued. The conventional weapons development program, revised in 1995, requires the development process of conventional weaponry to be divided into demonstration stage, program stage, project development stage, design stage and production stage. Space projects project phase division and planning (QJ3133-2001) puts forward that projects should fall into nine stages, namely, mission requirement analysis, feasibility demonstration, program design, project development, final design, trial production, batch production, using and disposal [2]. GB6992 specifies products' life cycle to be made up of concept and definition, design and development, manufacturing and installation, use and maintenance, as well as the disposal stage [3]. The implementation guide of life cycle cost (IEC60300-3-3) divides life cycle into several stages, namely, demonstration, design and development, production, installation, use and maintenance, as well as scrap stage [4].

2 Products' Lifetime Cycle Stages with the Main Tasks at Home and Abroad

2.1 Domestic Lifetime Cycle Stages with the Main Tasks

Lifetime cycle refers to the whole process, from comprehensive argument, scheme design, preliminary design, detailed design and development, to production/construction, use and maintenance and finally the decommissioning treatment. The key

point of each stage is different, which embodies the idea of quality control throughout the whole life cycle [5].

Only in special cases, will the product be reused. For general products, lifetime cycle only includes the following seven stages, as is shown in Fig. 1.

At the comprehensive demonstration stage, the main task is to implement necessary test, establish the goal, study the feasibility and initially determined the main technical index, overall scheme, research funding, development cycle and the guarantees. The completion of this phase is to pass the review.

At scheme design stage, the main task is to analyze and distribute functions, establish functional baseline, summarize the project with much importance and risks, carry out risk control work, and then the work on the reliability, maintainability and supportability of engineering. The finish line is to pass related review.

At preliminary design stage, the main task is to analyze subsystem functions, determine various technical indicators and the design requirements, design early prototype test and evaluation, make the procurement plan, etc. Passing the preliminary design review is the sign of completion.

At detailed design and development stage, the main task is to make the overall design, analyze and evaluate the key technology, performance, quality and risks. Draw up the trial production plan, build the trial production line, confirm the manufacturing process, the manpower, material resources as well as the cost, and then go on the manufacture and testing of various models. Passing the detailed design review marks the finish of this stage.

At production/construction stage, the main task is to efficiently achieve project objectives within the stipulated time and cost. Make construction control of various design parameters, and conduct a comprehensive inspection on the mass production condition and the stability of the quality, then make system level evaluation of the results.

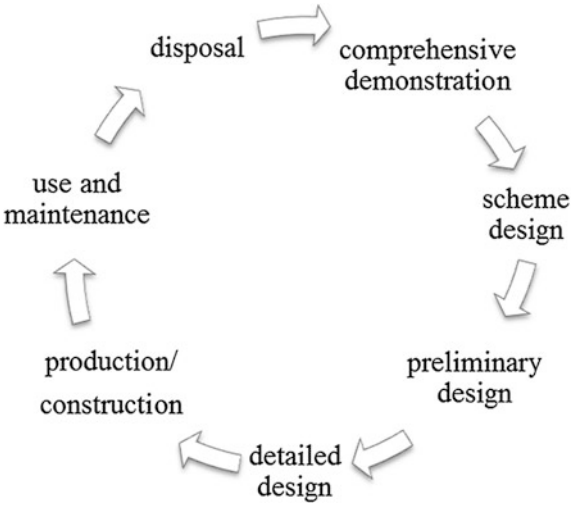


Fig. 1 Life cycle stages

At use and maintenance stage, the main task is to make system working in the user environment, and check regularly. Make maintenance scheme, analyze and evaluate the performance, quality, compatibility and interoperability. Besides, make field data collection and system evaluation.

At decommission stage, the main task is to use economical method to scrap or destroy the project with consideration of the impact on the environment, security and human health at the same time.

2.2 Life Cycle Stages and the Main Tasks of National Defence Project Abroad

The projects' stage division of all the countries in the world is almost the same with only slight difference of the name, management and the way to afford the cost. In general, there are two types: the American method and that of Western Europe.

According to America, the stages are concept, demonstration, development and production [6].

With respect to Europe, they are feasibility research, definition, development and production.

Now take American method as an example to analyze the main tasks of each stage.

(1) Concept stage. It includes three periods, namely, directed basic research, exploration and advanced technology development period. The first two periods equal to domestic advance research stage while the last one corresponding to the early development stage.
The target of directed basic research is to achieve the military goal through experiments and tests.
During the exploration period, some key technology with potential practical feasibility, such as new component or material, is under research and tests. The achievement of this stage is the prototype.
In the period of advanced technology development, the main tasks are to conduct computer simulation and calculation of advanced weapon system according to basic tactical requirements, and prove the feasibility of new technology, then develop the prototype after various tests.

(2) Demonstration stage. It mainly has two tasks. Firstly, analyze the use condition of the new technology and solve the problems left by feasibility research stage, put forward integrated tactical index of certain weapon type. Secondly, conduct demonstration, design and tests of the former research plan, which corresponds to the definition stage of European method. Then, within prescribed time, produce certain number of sample missiles to perform the test.

(3) Development and production stage. Some work on research and tests are still to be done. Later, conduct technical identification and combat use identification and tests, namely the type test of our country. But what is different is that

developed countries usually make massive research on the manufacturing technology, which reduces the cost and improve the quality.

The research cycle of overseas projects is almost the same with that of ours. But we need to study the recent application of the new technology which hasn't carried out previously research. In addition, domestic research stage includes the feasibility study and exploratory development in American concept stage, so the research and development work has to be done in the early development stage.

3 Specific Products' Life Cycle Stages with the Main Tasks

3.1 Weaponry's Life Cycle Stages with the Main Tasks

The development of weaponry is complicated system engineering. Different model has its unique characteristics because of the work, property and complexity which greatly influence life cycle stage division and the tasks. In addition, the system and condition of a country also have an impact on the stages and tasks. Chinese different weaponry life cycle division, together with that of America, is listed in Table A.1.

3.2 Aircrafts' Life Cycle Stages with the Main Tasks

Aircraft's life cycle starts from demonstration, through decommission and reclamation. In our country, it is regulated that aircraft's life cycle can be divided into stages of development, purchase, use, safeguard and decommission stage. Now we will introduce the research stage division method of several typical airlines abroad, and then sum up general division theory.

3.2.1 Boeing Company

The research stage can be separated into project definition, cost definition and production stage.

The main tasks of project definition stage are to build the construction baseline, make a schedule of the process and break down the structure.

At cost definition stage, the main work is to approve the configuration, carry out detailed design, purchase draft of the equipment and start marketing activities.

During production stage, it is supposed to complete the contract, issue configuration files and list of BFE in the engineering department, determine the equipment to be bought, the standard, material, tests, configuration definition, prototype, drawings and data, customer service information, etc.

3.2.2 Airbus

The development stage is generally divided into stages of feasibility study, concept, definition and development stage.

At feasibility study stage, the main work is for business observation, analysis of market situation and market demand, to determine the optimal potential plane concept.

At concept stage, the main work is to optimize the plane concept, form the aircraft configuration baseline, and elaborate it from aspects of product concept, market, production, maintenance. Then determine the technical requirements, business risk, division of work and cost, etc.

At definition stage, the main work is to determine the aircraft specification and business scheme, to design aircraft parts, raise funds to create production condition.

At development stage, the main work is to form the material bills and all needed data, finish preparation for manufacture, produce parts and assemble them together for system level tests and verification. Finally carry out complete factory test to achieve the goal of quality and reliability.

3.2.3 Bombardier

The research stage is usually developed into concept definition stage, preparation stage, preliminary/joint definition stage, detailed definition stage, release of product definition stage, product certification stage and product validation stage.

At concept definition stage, the main work is to carry on concept verification, feasibility study, configurations definition and other preliminary commercial activity.

At preparation stage, the main work is to approve the commercial operation, determine the basic structure and the aerodynamic mainline, define the system structure and choose the key suppliers of purchased parts.

At preliminary/joint definition stage, the main work is to make preliminary design, form freezing interface, analyze life cycle cost control and eventually finish the design review.

At detailed definition stage, the main work is to refine the preliminary design, complete the engineering layout drawing and organize the detailed design review.

At release of product definition stage, the main work is to confirm all the data needed for manufacture, assemble the system, freeze product design and complete product definition.

At product certification stage, the main work is to complete the product certification, prove airworthiness of the products.

At product validation stage, the main work is for the improvement of the product and delivery confirmation [7].

3.2.4 General Aircraft Life Cycle Stages

Although the above three stage division methods are different in start-stop node and the degree of detail, the overall content is basically the same. Generally, civilian aircraft life cycle includes the project demonstration stage, definition stage, development stage and batch production stage.

At project demonstration stage, the main work is to analyze market demand, predict development cycle, funding, risks and write project proposal.

At definition stage, the main work is to complete the overall research and system definition, select suppliers and sign contracts, conduct application for the airworthiness, go on preliminary design of main parts, manufacture the prototype and get all the data required for detailed design.

At development stage, the main work is to start detailed design, release a full set of production pattern, complete component and the final assembling, implement ground experiment and research, finish the aircraft production, test flight and get airworthiness qualification, model certification as well as production license, etc.

At batch production stage, the main work is making production plan according to the market demand, setting up and improving the sales and customer service system, improving the aircraft through the follow-up development and field data, making assurance of continued airworthiness, expanding the market, keeping the products' serial development.

3.3 Automobiles' Life Cycle Stages with the Main Tasks

The process of Car design development mainly includes the project plan, concept design, engineering design, prototype test and production starts.

At project plan stage, the main work is to conduct market research and identify model needs, analyze the project feasibility, including the policies, regulations as well as its resources and the R&D capabilities. Preliminarily set design goals for the new model, such as the body form, dynamic parameters, chassis assembly requirement and strength requirement, etc.

At concept design stage, the main work is to develop a detailed research plan, determine the time node in all stages of design, reasonably assign tasks, carry on the cost budget, make parts list form for the follow-up development work, put forward the layout requirement and characteristic parameters of the assembly and the parts, complete the overall layout sketches and design renderings, make clay model for the wind tunnel test, determine the final model and freeze it.

At engineering design stage, the main work is to refine the vehicle design, complete the design for various assembly and parts, determine its structure, characteristic parameters as well as quality requirements, coordinate contradictions between assemblies and vehicle, meet the requirements of the target performance. Finally confirm the vehicle design, prepare detailed product specifications and detailed spare parts list, verify the laws and regulations.

At prototype testing stage, there are two aspects: performance test and reliability test. Performance testing is used to verify if the assemblies can meet the design requirements, in order to find problems in time and make modification. Reliability test verifies the strength and durability. The test forms are ground test, road test, wind tunnel test and impact test, etc. According to the result of trial production and test, improve the design, and then turn to the second round production and test, until the performance of the new model is confirmed.

At production start-up stage, the main work is the preparation for the production process chain, all kinds of equipment, production line and so on. Among them, some should be started in the test stage, such as mold development and manufacture. Repeatedly perfect stamping, welding, painting and assembly production line. Start small batch production to further verify the reliability of the product. After 3 months without big problems, it is supposed to officially start the production.

3.4 Ships' Life Cycle Stages with the Main Tasks

The life cycle of the hull structure can be divided into six stages: preliminary design, detailed design, production design, construction, use and repair, scrap or dismantled [8–11].

At preliminary design stage, the main work is to determine the overall performance and main technical indexes of the ship and the principle of the system, identifying the basic technology form, working principle, main parameters, basic structure, equipment type of the product by theoretical calculation and necessary tests. It ends with the completion of hull specification and the cross section structure diagram.

At detailed design stage, the main work is to solve basic and critical technical problems by calculation and drawing, and eventually identifying all the technical performance, hull structure, materials, equipment, technical requirements and standards. It ends with the completion of typical structure diagram, shell expansion diagram, the rib line diagram, deck structure diagram, transverse bulkhead structure diagram, hull structure node diagram, welding specification tables, etc.

At production design stage, it is supposed to draw diagrams and charts with process requirement and production data, on the basis of the detailed design, combining with the shipyard process and production organization mode, consulting the process stage, construction area and the installation unit. It ends with the completion of the hull block division diagram, headroom layout diagram, sectional program diagram, BOM table, block test chart, segmented nesting figure, subsection processing map, etc.

At construction stage, it is programmed to test the safety, construction quality and the performance. Hull structure inspection items mainly concentrate on the construction process, the structural material and size, profile section, welding seam, coating, water tightness, structural integrity and the empty weight.

At use and maintenance stage, it is needed to conduct regular inspection and maintenance for the hull structure. Thickness is one of the most important inspection items comparing with others, because it has the most serious impact, and can be effectively assessed. Hull structure repair mainly solve the following problems: determine whether need to repair, repair method and scope, matters needing attention. Repair, inspection, evaluation and maintenance process cycle, appearing again and again.

Ship modification engineering is a major engineering project between ship repair and reconstruction. At this stage, there are many types of changes, such as the use, size, loading area, major equipment. Before the conversion, it is needed to know the requirements, involved convention and specification, and also the modified design drawing. After modification, complete documents and drawings should be given.

At scrap or dismantling stage, it provides a sustainable development and environmental friendly way for the retired ships. In May 2009, IMO adopted the "2009 Hong Kong convention on international security and harmless environmental demolition". It not only involves the whole process of ship design, construction, operation and dismantling, and also the disposal of harmful substances on the ship.

3.5 Constructions' Life Cycle Stages with the Main Tasks

As for constructions, life cycle refers to the whole process from idea to construction, until demolition. It can be divided into project demonstration, design, construction preparation, construction, completion acceptance, use and maintenance, scrap and removal [12–16]. Different stage has its unique work and key points.

At project demonstration stage, the main work is to clear customer demands, carry out land investigation, survey the condition of the location, conduct feasibility study and risk analysis, plan construction period and funds, then write the project proposal and feasibility report.

At design stage, the main work is to determine construction information such as the type, position, functional requirements. Then select a scheme, identify and solve technical problems, including structure design, professional drawings, etc.

At construction preparation stage, the main work is to arrange construction schedule, determine the node, and specify construction tasks, technical measures and quality requirements, considering the influence of season. Be familiar with drawings and complete preparation for the equipment, material, manpower, water, electricity, etc.

At construction stage, the main work is to collect the original technical data, archive the files, implement construction as it is ruled and regulated, supervise and control the quality, progress and cost, ensure safe construction conditions, conduct real-time inspection of security issues. What's more, environmental problems should also be solved by noise and waste control.

At completion acceptance stage, the main work is to clear accounts, finish the construction drawings. Conduct a comprehensive inspection acceptance of each

quality index, such as quality, time, safety, environment and others, ensuring satisfaction of the requirements and engineering standards.

At use and maintenance stage, it should strictly enforce quality review and maintenance of the construction. Repair when needed and guarantee the quality when in use [17].

At scrap and demolition stage, demolish the building as the plan, Reshuffle land use, classify the waste materials and recycle them, pay attention to the safety and environmental protection.

3.6 Other Projects' Life Cycle Stages with the Main Tasks

The whole life cycle of tunnel project is divided into the following four stages: project preparation stage [18], including project proposal, construction site selection, feasibility study, field trip, construction preparation; project construction stage, including construction, completion inspection and acceptance; project operation stage, including use, maintenance, property management; waste disposal stage, including scrap or dismantling process.

For the power distribution system and large water conservancy construction project, life cycle can be divided into design stage, construction stage, operation and maintenance stage. Design stage includes project design, project proposal, feasibility study, design bidding, preliminary design, technical design and construction drawing design [19, 20]. Construction stage includes construction and project acceptance. Operation and maintenance stage includes trial use, operation, maintenance and equipment update, until the last scrap.

In addition, the entire life cycle of the bridge project can be divided into the design, evaluation, construction, inspection, service, scrap and so on [21].

Highway and airport pavement [22] are through planning, analysis, design, construction, operation, maintenance and collapse/demolition stages.

4 Relationship of the Life Cycle Stages

In order to form a good mechanism of new products, it is not enough to break down the whole life cycle. It also needs analysis of the relationship between the stages.

Firstly, novel and practical conception is the basis and premise of the follow-up work. Comprehensive demonstration is an effective way to validate design concept. Plan designing is the start to concrete the abstract idea. Preliminary design is to refine the idea on the basis of the plan designing, and draw an outline for detailed design. Detailed design is the crystallization of previous work, and also provides build reliable guarantee for construction. Based on detailed design, production stage materializes the original idea for the first time. It is the target object of the maintenance phase, and can reflect the advantages and disadvantages of each stage, thus

providing guidance for further optimization and improvement. Disposal processing is projects' final destination, but the arriving time will be influenced by previous condition. Design flaws, errors, improper use or maintenance will delay the arrival of the last stage, while good idea, proper use and maintenance will prolong the using stage. It is the best results we hope to see.

Different stages have different tasks. It looks separate, but actually related closely. Any flaw will make failure arise in advance. Therefore, attention should be paid to every work of each stage and their mutual influence, especially the feedback information from later stages. Thus, problems could be found and solved in time, which reduces the risk of loss, shortens development cycle, and strengthens competitiveness.

5 Conclusion

Systems engineering is the science of organizing and managing engineering system, and also is the engineering technology to solve engineering problems of the whole process. Full life cycle process is the elaboration of system engineering in time dimension, and also the focus of engineering optimization. Stage division is the necessary means to carry out life cycle theory, and also is a prerequisite for efficient work distribution.

By stage division, it is convenient to conduct life cycle management, to timely discover and maintain the flaw, which avoids rework, increasing products' one-time pass rate, saving the resources and costs, greatly shortening the time of development and optimization.

In order to help readers master the life cycle phase partitioning method and effectively arrange tasks, this chapter provides not only domestic and foreign stage division method with a simple comparison, but also analyzes the stage division of different projects, such as the weaponry aircraft, automobile, ship and some others. Apparently, they have differences, but generally begin with comprehensive demonstration, then go through scheme design, preliminary design, detailed design, production/construction, use and maintenance, eventually end in retirement. The focus of each stage is quite different, but they are closely linked, greatly influenced by each other. The former stage is the foundation of the latter, while the latter stage is the verification of the former.

In general, it is of great significance to have overall understanding of the project/product at the beginning, and put the theory of system engineering and lifetime cycle into engineering practice, making them serve modern enterprises and help produce products with high reliability, long life and low cost.

Appendix

Table A.1 Comparison between Chinese and American weaponry life cycle stage division

Nation	China				America		
Weaponry type / Stage division	Ground-to-ground missile	Winged missile	Surface-to-air missile	Aerospace project	Strategic missile	Missile nuclear weapon	Aerospace project
Stage 1	Mission requirement analysis				Research	Scheme design	O
Stage 2	Feasibility demonstration				Exploring development	Feasibility study	A
Stage 3	Scheme Design					Engineering development	B
Stage 4	Prototype	Prototype	Independent circuit play and ground equipment Prototype	Prototype	Advanced development		
Stage 5	Sample	Sample	Closed loop play and ground equipment sample	Sample	Engineering development	Manufacturing	C
Stage 6							D
Stage 7	Finalizing design		On-orbit test		Manufacturing	First production batch	E
Stage 8	Trial production						
Stage 9	Batch production					Mass production and storage	
Stage 10	Use and improvement						
Stage 11	Disposal					Decommission and disposal	

References

1. Kang R, Li RY Basic tutorial of project system engineering. Beijing: Beihang University
2. Mao JR, Xu Y (2009) Development stage division analysis of aerospace model. Aerosp Ind Manag 9:008
3. Li ZH, Wang HS (2003) Several problems of high-speed train maintenance. Railw Locomotive Car 23:Suppl 2
4. Liu WJ, Liu LJ (2012) Electric locomotive LCC model research and application. Mech Manag Dev 5:95–98
5. Huang SX, Fan YS (2004) Product life cycle management research review. J Comput Integr Manuf Syst 10:1
6. Yu YX (1994) Discussion about the development stage and the goal. J Aero Weapon 2:6
7. Li Y, Zheng SF (2008) Several problems of civil aircraft research stage division. J Aeronaut Stand Quality 3:8
8. Chen DY, Wu ZJ (2005) General principles for repairing defects of the ship structure. China Shiprepair 6:28–30
9. Lin H (2005) Analysis of vessel conversion type and major points of converting technology. China Shiprepair 2:15–17
10. Liu XT (2008) New ship maintenance management. Marine Technol 2:47–50
11. Wang SL, Liu YD (2000) The theory of ship design. Dalian Industry Press, Dalian
12. Fu SS, Ji HJ (2009) Preliminary study of real estate projects' LCC stage control. Mod Sci 5:105
13. Gao YB (2008) Sustainable residence research based on theory of whole life cycle. Archit J11:44–46
14. Han Y, Cheng H (2010) Framework study of engineering whole life cycle design. Sci Technol Prog Policy 27:19
15. Wei YH (2012) Optimization analysis of X project construction, organization and design. South China University of Technology, Guangzhou
16. Zheng L (2006) Interface management of real estate project. China Acad J Electron Publ House 1:73–75
17. Huang XB (2008) Life cycle control research of construction project. Jiangxi University of Science and Technology
18. Dong YK, Han B, Guo JT (2008) Tunnel project cost control research based on lifetime cycle. West Explor Eng 20:8
19. Chen J, Yin ZY, Li ZX (2008) Preliminary study of the large water conservancy construction project's total life cycle management. Water Resour Dev Res 7:6
20. Su HF (2012) The theory and study method of power distribution system life cycle management. North China Electric Power University, China
21. Peng JX (2008) Bridge lifetime design method research based on LCC. Hunan University, Hunan
22. Xu Y, Hu GL (2009) Lifetime cycle analysis of airport pavement. Henan Build Mater 1:111–112

Calculation of Failure Rate of Semiconductor Devices Based on Mechanism Consistency

Cui Ye, Ying Chen and Rui Kang

Abstract Failure rate data of components is widely used in reliability design and analysis, such as reliability apportionment, reliability prediction, FMECA and so on. Currently, there are two major ways to evaluate failure rate: using prediction handbooks and accelerated life test. Updates to data in handbooks always are delayed, while accelerated life test costs lots of time and money, especially large-scale integrated circuits are expensive. In view of the defect, using corresponding test data provided by manufacturer's website to calculate failure rate is proposed based on failure mechanism consistency. Test data is analyzed and calculated to get failure rate of Semiconductor devices with each influencing mechanism, and values of failure rate under each mechanism are added together to get total failure rate based on the assumption of competition between mechanism.

1 Introduction

Failure rate of component is needed in reliability design and analysis, such as reliability apportionment, prediction, test, FMECA and so on. In electronic devices, the key components to achieve functions are mostly semiconductor devices. So, accurately estimating the failure rate of semiconductor devices is basis of proceeding reliability analysis accurately and effectively.

Now there are two ways commonly used in engineering to assess the failure rate of semiconductor devices. One is using prediction handbooks, such as MIL-STD-217F or GJB299C, which providing component failure rate data. Data update lag in handbooks cannot reflect the current semiconductor device design and manufacturing level very well. The other way is conducting accelerated life test, which accelerating components to fail by increasing the stress on the premise of the failure mechanism unchanged, then using model to calculate component life under practical conditions.

C. Ye (✉) · Y. Chen · R. Kang
School of Reliability and Systems Engineering, Beihang University, Beijing, China
e-mail: buaahhz@163.com

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_153

Many scholars have done the research for accelerated life test [1] studied test stress type selection, determination of test conditions and so on [2, 3] studied accelerated test failure models. These studies have important theoretical and engineering value, and play a catalytic role for accelerated life applications in the field of reliability.

When the sample size is large, data obtained from accelerated life test is more accurate than using handbook. However, if we use accelerated life test to assess reliability of electronic products, accelerated test needed to be implemented for every component with long time and high cost, especially high cost of LSI chips.

Based on failure mechanism, this chapter provides a way to assess failure rate of semiconductor devices by analyzing existing experimental data under corresponding mechanism.

2 Related Theories

2.1 Failure Mechanism and Mechanism Consistency

Semiconductor devices mainly include wafer and package two parts in structure, so the failure rate can be divided into wafer failure rate and package failure rate to calculate. According to current research analysis, there are two types of component failure: time-related degradation and accidental failure. Due to impurities, process control and other factors, chip manufacturing defects will be caused in wafer fabrication, oxidation and photolithography process of semiconductor devices. These defects can induce various mechanism, such as metal key compounds, stress migration of metal wire inside the chip and so on, which is the root of dispersion. Studies [4, 5] show that semiconductor wafer failure can be considered to obey the exponential distribution in engineering, and the main environmental factor of affecting this failure is temperature. After the chip manufacturing, semiconductor devices are packaged to form complete devices. The main mechanism of package failure is crack propagation of lead bound interconnected parts, and its main factor is temperature cycling. Package failure is usually considered to obey two-parameter weibull distribution in engineering.

Accelerated life test should be taken in the premise that failure mechanism remains unchanged, that is to say, failure mechanism in accelerate tests is consistent with failure mechanism in field experiments. Experimental data is used to obtain the corresponding characteristic parameters of lifetime distribution, and then we can calculate the characteristic parameters of the product life distribution in true stress conditions according to the relationship between stress and life. Consistency between accelerated test data and field test data is important criteria of whether accelerated testing is successful. The key to implement test successfully is to explore the failure mechanism and its impact on product. Found in study, main failure mechanism of semiconductor device wafer is intermetallic compounds and metal wire stress migration, and the main environmental stress of affecting failure is steady temperature. Mechanism of semiconductor packaging is wire fatigue fracture

and the main environmental stress affecting the failure is temperature cycling. So semiconductor devices should be conducted accelerated test respectively with temperature and temperature cycling stress aiming at different mechanism.

2.2 Competition of Mechanism

For semiconductor devices having several kinds of mechanism, it is assumed that every failure mechanism is independent of each other when calculating characteristic parameters of the device lifetime distribution. Independent failure mechanism is not resulted from other mechanism in system but from the interaction between environmental conditions and internal factors. Every mechanism competes with each other, so series model is chosen to be reliability model of semiconductor devices.

Since every failure mechanism is resulted from the interaction between environmental conditions and internal factors, the impact of environmental conditions on the failure is reflected in the life distribution under the corresponding mechanism. We use accelerated test failure model which involves stress in test and reality to calculate acceleration factor, then calculate characteristic parameters in real life distribution combining with accelerated test data.

3 Approach to Analyze Test Data

According to the theory of failure mechanism consistency, HTOL test should be conducted for semiconductor device wafer failure, while temperature cycle test should be conducted for package failure. If failure rate of a variety of semiconductor devices is assessed by designing and implementing accelerated life test, it would cost a lot of money and time. At present, many large manufacturers make public a large number of experimental data. So these data could be screened according to the mechanism consistency. Then failure rate under the mechanism can be obtained by analyzing and calculating these data. For semiconductor devices, we can screen data to get appropriate data of the HTOL test and temperature cycling test to calculate failure rate of wafer failure and package failure. This chapter uses the reliability test database in ADI as an example to illustrate the way to assess semiconductor device failure rate by using existing reliability data.

3.1 Calculation of Failure Rate of Wafer

It is assumed that large electronic component manufacturers, such as Analog Devices, NS, AMD, hp, MAX, Texas, Toshiba, have considerable level of component control process during the same period, and wafers manufacturing with the

same process have consistent dispersion. For component manufactured in ADI, its wafer failure rate under specified operating temperature can be obtained directly from ADI reliability database. Analogy to similar products based on information of package, process and parameter could be used to assess failure data of component from other manufactures which is similar to component in ADI database.

Semiconductor wafer failure is considered to obey the exponential distribution. In no replacement censoring life test, lower confidence limit of average life product obeying exponential distribution in the confidence level of $1 - \alpha$ is:

$$\theta_L = \frac{2T}{\chi^2_{1-\alpha}(2r+2)} \tag{1}$$

In the equation, θ_L is the lower confidence limit of average life; χ^2 is Chi-square distribution, Value depending on the number of failures and confidence; T is total test time.

ADI's website provides data with no failure of censored time HTOL test. It is assumed that N is number of HTOL test samples; H is HTOL test duration and A_t is acceleration factor. So equivalent total test time is $T = N \cdot H \cdot At$, and χ^2 values $\chi^2_{1-\alpha}(2)$, for r = 0, confidence level is $1 - \alpha$. According to the Eq. (1), it can be launched expected failure rate of device wafer is:

$$\lambda = \frac{\chi^2_{1-\alpha}(2r+2)}{2T} = \frac{\chi^2_{1-\alpha}(2)}{2N \cdot H \cdot At} \tag{2}$$

In the equation, λ is expected failure rate of device wafer.

Addition to knowing the number of test samples, test time and the number of failure data, it is necessary to obtain the value of acceleration factor At in HTOL test to estimate expected failure rate of device wafer at a certain confidence level. Acceleration factor is calculated based on accelerated testing and field trials stress levels, describing the relationship between environmental stress and life distribution parameters.

3.2 Calculation of Failure Rate of Package

Semiconductor device forms a complete device through package after manufacturing chip. The main mechanism of package failure is crack propagation of lead bound interconnected parts, and the main factor affect the failure is temperature cycling. Package failure is usually considered to obey two-parameter weibull distribution in engineering. The key to calculate package failure rate is the calculation of package shape parameter β and scale parameter η.

ADI provides a wealth of temperature cycling test data of components in a variety of processes. Cycling condition is −65/+150 °C, etc. these reliability tests

are censored time, almost no failures. The approach to obtain β and η in failure probability density distribution by using these on failure data is described in the following.

3.2.1 Shape Parameter β

Currently, shape parameter data in fatigue Weibull distribution is already accumulated in engineering. Study [6] points that according to research from aviation sectors of United States, Britain, Japan and Australia and other countries, the limit of shape parameter in Weibull distribution which fatigue crack formation and propagation obeys for metal structure is approximately 2.2. Boeing's statistical analysis from lots of test data in Weibull distribution also show that, shape parameter in Weibull distribution of various metal structure fatigue fracture is in the range of about 2.2–4.0. In this chapter, the limit of the shape parameter β in Weibull distribution of crack propagation in lead bound interconnected parts is 2.2.

3.2.2 Scale Parameter η

After determining the shape parameter β based on engineering experience and historical data, test data with no failure is used to obtain product's characteristic life η. When the shape parameter is known, one-sided confidence limit of characteristic life η in Weibull distribution at confidence level of $1 - \alpha$ can be calculated by using the following equation:

$$\eta_L = \left[\frac{\sum_{i=1}^{n} t_i^{\beta}}{-\ln \alpha} \right]^{\frac{1}{\beta}} \tag{3}$$

Scale parameter η in Weibull distribution under different temperature cycling is different. As the changes of scale parameter η in Weibull distribution of wire fatigue fracture caused by different temperature cycling consistent with the inverse power-law relationship, Characteristic life η in real case can be calculated according to the inverse power-law relationship.

According the obtained shape parameter β and scale parameter η, package failure rate function can be achieved.

$$\lambda(t) = \frac{\beta}{\eta^{\beta}} t^{\beta - 1} \tag{4}$$

Then combined with the device operating time, the value of package failure rate can is worked out. Finally, failure rate of wafer and package are added to calculate the failure rate of semiconductor devices.

4 Calculation Example

In this section, Field effect transistor IRF5210 will serve as an example to describe the calculation method for failure rate of semiconductor devices. The package of IRF5210 is TO-220AB, and manufacturing process of its wafer is >2.5 μm^2 Bipolar.

Firstly, select its production process in the ADI database [7], and enter the working temperature 70 °C, then the failure rate of IRF5210 wafer at 70 °C can be given directly by running in the background as shown in Fig. 1.

The specific calculation method is calculating the acceleration factor A_t by using Arrhenius model based on the operating temperature and the test temperature. After knowing the acceleration factor A_t, failure rate of corresponding wafer could be work out by the Eq. 1 with appropriate HTOL test data in ADI database.

Arrhenius model equation is shown below:

$$A_t = \exp\left[\frac{E_a}{B}\left(\frac{1}{T_u} - \frac{1}{T_s}\right)\right] \tag{5}$$

In the equation, E_a is the activation energy; B is Boltzmann constant, value 0.00008623 eV/K; T_u is operating temperature; T_s is test temperature.

Then collect data with TO package and process >2.5 μm^2 Bipolar in temperature cycling test data provided in ADI's database, shown in Table 1.

Use Eq. (3) to calculate respectively the point estimates of the η under two different cyclic temperature. Generally, we choose confidence limit under the level "$1 - \alpha = 0.5$" as point estimate of characteristic life η. Then we calculate K and n based on the inverse power law equation combining with different temperature difference in cycling test and corresponding estimate of η. the inverse power law is shown in the following.

Fig. 1 Page of calculation of failure rate of wafer

Wafer Fabrication Data

Process Technology: >2.5um^2 Bipolar

>2.5um^2 Bipolar Life Test Data Summary	
Overall Sample Size	20028
Qty. Fail	0
Equivalent Device Hrs. @ 70 deg C)	763356016
FIT Rate (60% CL, 70 deg C)	1.2
MTTF (60% CL, 70 deg C)	833094343
FIT Rate (90% CL, 70 deg C)	3.02
MTTF (90% CL, 70 deg C)	331522039
Calculations assumes 0.7 eV Activation Energy	
To recalculate the summary table with a different operating temperature, enter a temperature and click Recalculate: 70 °C Recalculate	

Table 1 Cycling test data with TO package and process >2.5 μm^2 bipolar

Temperature cycling (°C)	Test time/h	Sample size	Number of failure
−65/+150	500	1,855	0
−65/+150	1,000	746	0
−40/+125	500	2,645	0
−40/+125	1,000	1,038	0

$$\eta = \frac{1}{K(\Delta T)^n} \tag{6}$$

After calculating K and n, we calculate η in the real working condition −40/70 °C by using the inverse power law. Then work out that failure rate of IRF5210 package is 4.2×10^{-9}/h based on Eq. (4) with working time 10 h.

Based on mechanism competition model, reliability model of semiconductor device is series model. So, failure rate of IRF5210 is obtained by add failure rate of wafer and package together. Finally, we calculate that failure rate of IRF5210 working 10 h is 7.2×10^{-9}/h.

5 Comparison with Failure Rate Assessment Method Based on Handbook

From the above results, failure rate of IRF5210 is conservative calculated as 7.2×10^{-9}/h based on test data provided by component manufacturers. While it is calculated as 1.62×10^{-8}/h through calculation method provided in Appendix A.2.3.2 of GJB299C.

The latest version of GJB/Z299C-2006 "Electronic Equipment Reliability Prediction Handbook" was released in 2006. In the past 7 years, the manufacturing process and quality of components improved and the failure rate significantly reduced. So, failure rate data in GJB299C might be too large to reflect the current level of component manufacturing.

Reliability prediction in database provided by Manufactures is based on test data, these data in database has been kept up to date. Therefore, these data can reflect the current level of design and manufacture of components. For example, data in ADI' website last updated on August 06, 2013. This method using the latest reliability data to calculate failure rate can be used engineering application.

6 Conclusion

This chapter first describes the major semiconductor device failure mechanisms; and then uses corresponding reliability test data provided by semiconductor device manufacturers to calculate failure rate of component under various mechanism based on mechanism consistency; finally, add failure rate of different mechanism together to get failure rate of semiconductor devices under the assumption that the mechanism competitive relationship.

Using these test data not only can save cost of test, and can reflect the current component design and manufacturing better than handbooks because these data have been kept up to date. The method calculating failure rate introduced in this chapter combines reliability data analysis knowledge with test data. If components have sufficient test data, this approach can be used in reliability prediction.

References

1. Zhang Q, Liu C (2011) Reliability accelerated testing(RAT) for electronic equipments. Electro-Opt Technol Appl 26(4):81
2. Mao S (2003) Acceleration model in accelerated life testing. Qual Reliab 2:15–17
3. Li X, Jiang T (2007) Review of multiple-stress models in accelerated life testing. Syst Eng Electron 29(5):828–831
4. Moon J-F (2006) Time-varying failure rate extraction in electric power distribution equipment. In: 9th international conference on probabilistic methods applied to power systems. KTH, Stockholm, Sweden
5. Joseph B, Moshe G (2006) Electronic circuit reliability modeling. Microelectron Reliab 46:1957–1979
6. Gao Z (1986) Fatigue applied statistics. National Defense Industry Press, Beijing
7. Reliability test database in ADI (2013) http://www.analog.com/en/quality-and-reliability/reliability-data/content/index.html. Accessed 8 Aug 2013

Legal Aspects of Engineering Asset Management

Joe Amadi-Echendu and Anthea Amadi-Echendu

Abstract Immovable assets include engineered infrastructure such as buildings, manufacturing plant, roads and railways. These assets are built on land. In most jurisdictions, proposals to acquire and/or establish immovable assets on landed property must comply with several legislative provisions. In many instances, the acquisition and/or establishment of an immovable asset becomes embroiled in legal disputes between contending stakeholders asserting rights to ownership/custodianship of land. Such disputes add to the costs of capital development projects, mergers and acquisitions, and influence decisions as to where an asset intensive business venture can be located. Legislation not only provides the means to resolve ownership/custodianship rights but also, it stipulates legal imperatives for control and utilization of engineering assets on landed property. It is in this regard that this chapter presents some of the legal aspects of engineering asset management.

Keywords Immovable engineering assets · Asset ownership · Legal imperatives

1 Introduction

When engineering structures such as buildings and process plant are built on land, they are regarded as immovable or landed property assets. In most jurisdictions, land is treated as a natural asset, and in indigenous areas in particular, ownership/custodianship and exploitation of land are often vested in hereditary alienable

J. Amadi-Echendu (✉)
Department of Engineering and Technology Management, University of Pretoria, Pretoria, South Africa
e-mail: joe.amadi-echendu@up.ac.za

A. Amadi-Echendu
Department of Business Management, University of South Africa, Pretoria, South Africa
e-mail: amadiap@unisa.ac.za

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3_154

rights. By convention, the creation and establishment of an engineering structure such as a building, a road, a manufacturing or processing plant on a piece of land is regarded as *developing* the land or making the land more valuable.

In the first instance, land has to be acquired before it can be *developed*, and the process of acquiring land usually confers title and establishes the rights to ownership, custodianship and exploitation. In the second instance, permission has to be obtained in order to operate any engineering asset built on land. The implication then is that an engineering asset such as a rail line, an airport, a manufacturing or processing plant also has to be *registered*, so that the registration process similarly confers certain rights to ownership, custodianship, control and utilisation.

Most organisations require engineering assets that are built on or erected on landed property in order to conduct business, and the process of acquiring and/or establishing such assets often becomes embroiled in legal disputes between contending stakeholders of the landed property. Such disputes have taken on new significance in the modern era of globalisation and sustainability imperatives. The prominence of environmental impact assessments is a case in point. There are many legislative directives in every jurisdiction (see, for example, reference [1]), and a key feature of legislation stipulates that consultation with the stakeholders is a primary requirement for landed property development projects.

Both practitioners and scholars would readily acknowledge that engineering asset management commences with the acquisition of an asset but, often times, the processes involved in acquiring engineering assets must comply with a number of legal regulations and legislative provisions. The erection of a shopping mall, the construction of a road or railway line, the development of an airport infrastructure, the establishment a manufacturing or process plant can easily be stuck in legal disputes as to ownership or custodianship of the landed property. At least, environmental legislation contains strict regulations and many business organisations have to overcome compliance challenges in order to establish, operate and maintain their engineering assets located within a jurisdiction.

For engineering asset managers who wish to limit their scope to already established infrastructure, manufacturing or process plant, the process of acquiring associated equipment, machinery and spare parts also involve legalities which are latent in the clauses contained in contract documents between supplier/vendor (the seller) and user client (the buyer). Warranty clauses often restrict the actions that can be taken by operators and maintainers of certain equipment and machinery. With the emergence of the so-called product servitization (cf: [2,3]), some original equipment manufacturers have become very strict with regard to the intangible aspects of the engineering assets they produce. Trademarks and unique design features need to be respected by suppliers, vendors and users of equipment as original manufacturers seek to appropriate annuity income from embedded technological innovation. Companies that build and construct engineering infrastructure attach their logo or insignia to suggest or indicate some legal right of disclaimer, indemnity, preference, or protection from competitors.

As illustrated in Fig. 1, the management of an asset, particularly an engineering asset built or erected on landed property, involves a number of phases (i.e., acquire,

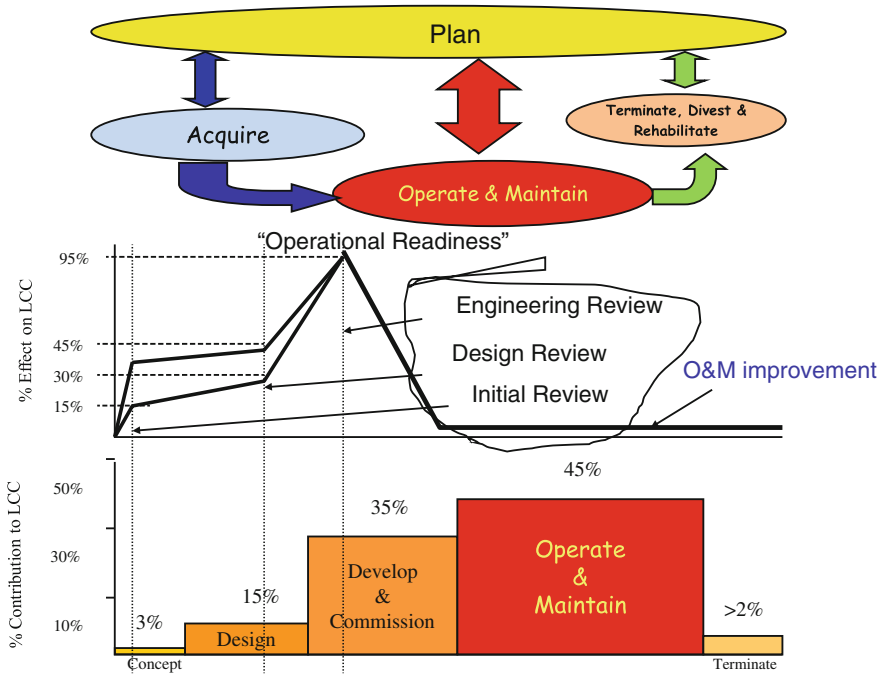


Fig. 1 Engineering asset management life-cycle phases and stages

operate and maintain, and divest), and life-cycle stages (i.e., concept, design, develop and commission, and terminate). The graphs in Fig. 1 depict that 95 % of the total life cost is typically determined during the acquisition phase of an engineering asset, hence, legal considerations that affect the acquisition processes are pertinent.

Although there are legislative influences at each phase or stage, however, this chapter focuses on the acquisition phase of an engineering asset. Legally, it is important to consider who

i. owns or has the rights to, and
ii. controls the use of an asset.

Psychologically, ownership/custodianship engenders a positive attitude to asset management but, ownership/custodianship is conferred as a legal right. Thus, it is pertinent to examine legislation that influence the *conveyancing* processes of conferring rights to ownership, custodianship or control of the landed property upon which an engineering asset base is established.

2 Establishing Legal Rights to Engineering Assets

The transactions and activities that result in the assignment of title deeds of landed property to an owner fall under the term conveyancing. According to [4], conveyancing involves business processes traditionally designed to transfer a landed property from one owner/custodian to another. In civil society and in general, ownership/custodianship is defined in terms of legal right, hence conveyancing is perceived as a legal convention. The reality is that the actual transactions that result in the transfer of ownership/custodianship of landed property are derived from activities of the various private firms and public agencies (role players as shown in Fig. 2) that are involved in conveyancing. The process of acquiring, establishing or transferring ownership/custodianship/control of an engineering asset also involves many business-to-business transactions. Public-private partnership developments of engineering assets typically involve legal rights in the form of concessions. Business-to-business mergers and acquisitions also involve legal rights to ownership, custodianship, and control of the associated engineering assets.

A significant ramification is that conveyancing processes are not limited within the conventional firm-level business management parlance, especially because the transactions and activities do not occur exclusively within a firm. Rather, during the acquisition phase, the transactions and activities that establish the rights to

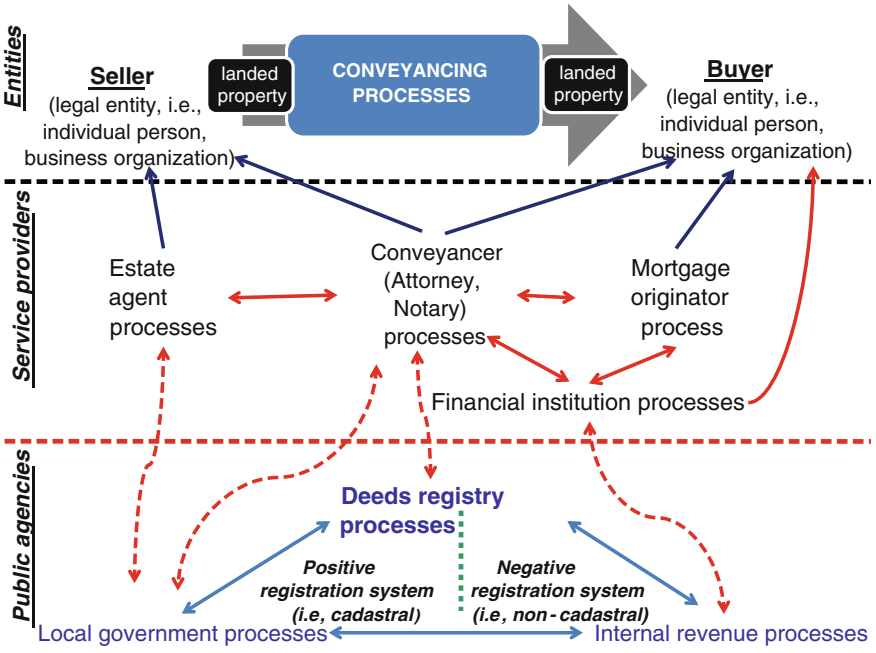


Fig. 2 A conceptual mapping of conveyancing processes

ownership/custodianship of an asset are *institutionally* carried out by different role players as depicted in Fig. 2, and these include private firms like banks and attorneys, as well as public agencies like the municipality, Master of Court, and Deeds Office. A subsequent ramification is that conveyancing includes wider legal influences on the processes of acquiring, for example, a building for business operations, or land on which a manufacturing plant is located. From a strategic asset management viewpoint, these legal aspects are not only significant during mergers, acquisitions, and divestments of engineering assets but also, they contribute significantly to the costs of such business transactions. At the tactical level, such legalities impose asset management inconveniences and overheads which operators and maintainers of equipment and machinery tend to relegate to infrastructure and estate managers.
Conveyancing processes broadly include:

i. valuation of property (e.g., by actuarial scientists, quantity surveyors, real estate agents)
ii. financing activities (e.g., by banks, financial institutions)
iii. contracts (e.g., by attorneys, notaries, and conveyancers)
iv. statutory registration (e.g., by local government, internal revenue), and of course,
v. custodians and owners (i.e., sellers and buyers).

Similarly, engineering asset management also involves:

i. valuation of an asset
ii. financing the acquisition, operations, maintenance and divestment of an asset
iii. contracts (e.g., supply chain, service delivery)
iv. registration of assets (for fiduciary and technical integrity compliance)
v. ownership and custodianship (legal responsibility).

The implication here is that conveyancing processes are not only necessary for asset management but also, they mirror activities which are repeated throughout the life cycle phases and stages of an engineering asset. It is in this regard that we identify and briefly examine some legislation that may influence the processes of acquiring, developing or utilisation of engineering assets built or erected on landed property in our case study country—South Africa.

3 Legislation Regarding Landed Asset Processes

In South Africa, conveyancing processes have evolved from indigenous land tenure culture and colonial interventions, through several formalised legal provisions and regulations (see, for example, [5]), to the current structure as illustrated in Fig. 3. The conveyancing processes essentially link the seller and the buyer, irrespective of whether the seller or buyer is a business organisation or an individual person. The process starts with an initial binding agreement between a buyer and a seller, and

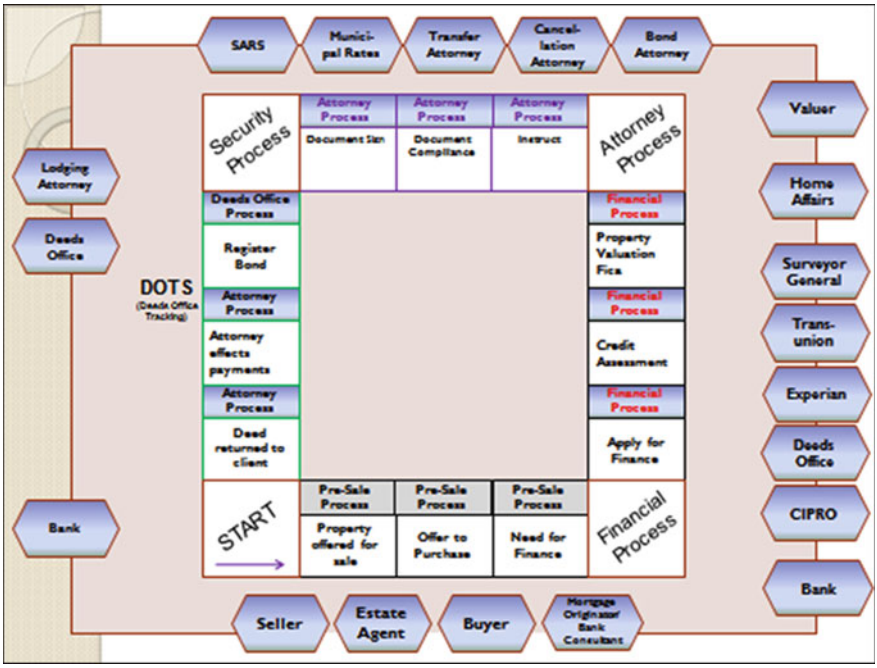


Fig. 3 A mapping of conveyancing processes in South Africa (see [4])

progresses through financial, legal, fiduciary and security scrutiny. The end result is a legal transfer of ownership and associated rights from seller to buyer. The structure of conveyancing processes in four other countries also involves similar role players and transactions, even though the actual flow of the processes may be different [4].

Where the buyer and seller are firms, the legalities increase in complexity depending on the legal form of the respective business organisations involved. That is, the number and precedence of legislation which must be considered or complied with depends on whether the seller or buyer, or both are limited liability companies, partnerships or simple sole traders.

For brevity, in South Africa, land registration is based on statute, typified by the Alienation of Land Act 68 of 1981 and Deeds Registries Act 47 of 1937. South African law defines real estate as "…immovable property which includes land, a building annexed to land, … and an object which is attached to immovable property where the attachment is so secure that separation would cause significant damage to either the immovable property or the land" [6]. Engineering structures without foundations on land are not deemed to be immovable property. Table 1 provides a snapshot of some legislation pertinent to the processes of acquiring landed assets in South Africa.

Table 1 Outline of some South African legislation regarding acquisition of landed property assets

	Legislation	Provision
a	The Upgrading of Land Tenure Rights Act, 1991 (Act 112 of 1991)	"…provides for the upgrading of various forms of tenure to ownership…"
b	The Restitution of Land Rights Act, 1994 (Act 22 of 1994), as amended: The Restitution of Land Rights Amendment Act 48 of 2003	"…provides for the restitution of land or equitable redress … as a result of past racially discriminatory laws or practices…"
c	The Deeds Registries Act, 1937 (Act 47 of 1937); The Sectional Titles Act, 1986 (Act 95 of 1986); The Land Survey Act, 1997 (Act 8 of 1997)	"…administration of the land registration … and rights; … regulates the survey of land, registration and administration of sectional title schemes…"
d	The Development Facilitation Act, 1995 (Act 67 of 1995)	"…a transparent and democratic … land use management system … to facilitate … development programmes and projects in relation to land"
e	Physical Planning Act, 1991 (Act 125 of 1991)	"…orderly physical development of…"
f	Matrimonial Property Act 88 of 1984; Marriage and Matrimonial Property Law Amendment Act 3 of 1988	"…interpretation, management and dissolution of property in terms of various marriage regimes…"
g	Insolvency Amendment Act 122 of 1993	"…regulate the registration of transactions in respect of immovable property after the expiry of caveats; … recovery of the value of immovable property disposed of unlawfully…"
h	Alienation of Land Act 68 of 1981	"To regulate the alienation of land in certain circumstances and to provide for matters connected therewith"
i	Public Finances Management Act No.1 of 1999 (PFMA)	"To regulate financial management … to ensure that all revenue, expenditure, assets and liabilities … are managed efficiently and effectively…"
j	Government Immovable Asset Management Act 2006	"To provide for a uniform framework for the management of an immovable asset … to ensure coordination of the use of an immovable asset with the service delivery objectives…"

4 Concluding Remarks

Conclusions should state concisely the most important propositions of the chapter as well as the author's views of the practical implications or consequences.

A remarkable feature implied in the sample legislation listed in Table 1 is the requirement to *register* an asset, either

i. for ownership/custodianship title, or
ii. for rights to control, exploit and utilise, or

iii. for fiduciary responsibility, or
iv. all the above.

Engineering asset management involves all activities necessary to ensure that an asset provides the means for the realisation of the profile of values desired by the associated stakeholders. A major assertion of this chapter is that engineering asset management activities have a legal basis. What this means is that, it is not only a cognitive preference (see [7]) but also, it is a legal imperative for an engineering asset manager to think and act as an owner of the asset. The audit trail implicit in conveyancing transactions further indicate that engineering assets have to be managed in a manner that demonstrates fiduciary compliance.

Whereas the sample legislation listed as *a* to *h* in Table 1 mostly deliberate on the landed property title and registration issues, legislation *i* and *j*, albeit particularly applicable to the public sector, stipulate efficient and effective management of assets. Infact, as shown in Fig. 4, the South African National Treasury Asset Management guideline arising from the PFMA (i.e., legislation ***i*** in Table 1) provides a detailed stipulation of data and information that should be contained in an asset register. From the preceding discussions, registration is probably one of the final steps in the acquisition phase of an asset, it is a legal requirement, and the template in Fig. 4 shows how to comply with the PFMA legislation. Although this template has been primarily designed for the registration of public sector controlled assets, it is also applicable for the registration of engineering assets controlled by private firms.

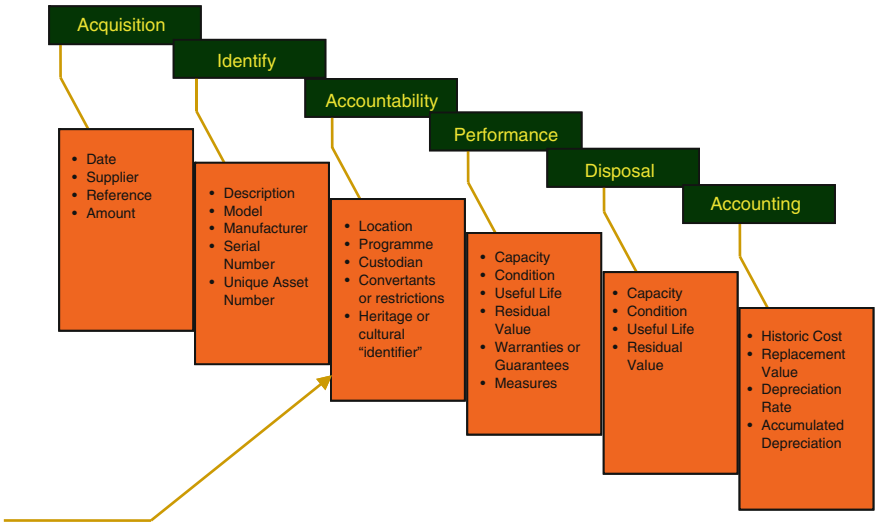


Fig. 4 Asset registration template (*Source* [8])

References

1. European Union (2013) Environmental impact assessment of projects—rulings of the court of justice. http://ec.europa.eu/environment/eia/eia_case_law.pdf. Accessed 5 July 2013
2. Neely A, Benedettini O, Visnic I (2011) The servitization of manufacturing: further evidence. 18th European operations management association conference, Cambridge, July 2011
3. Steunebrink G (2012) The servitization of product-oriented companies. MBA thesis, University of Twente, The Netherlands, 31 Aug 2012
4. Amadi-Echendu AP, Pellisier R (2013) Lessons for South Africa from the international *e*-conveyancing environment. Global business and technology association's 15th annual international conference. Henlsinki, Finland, 2–6 July 2013
5. Hodson TA (ibid) South African land tenure, past and present—a country report. http://www.spatial.maine.edu/~onsrud/Landtenure/CountryReport/SouthAfrica.html. Accessed 5 July 2013
6. Naumann U, Shawe L (2012) Multi-jurisdictional guide. Bowman Gilfillan Inc., Johannesburg. www.practicallaw.com/2-384-6156
7. Amadi-Echendu JE (2010) Behavioural preferences for engineering asset management. In: Concepts, definitions and scope of engineering asset management. Engineering asset management review, vol 1. Springer, Heidelberg. E-ISBN:978-1-84996-178-3
8. Asset Management Guideline—Framework for the recognition of assets. Version 3.2 April 2004 South African National Treasury

Author Index

P.W. Tse et al. (eds.), *Engineering Asset Management - Systems, Professional Practices and Certification*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-09507-3