Lecture Notes in Mechanical Engineering

Uday Kumar
Alireza Ahmadi
Ajit Kumar Verma
Prabhakar Varde *Editors*

Current Trends in Reliability, Availability, Maintainability and Safety

An Industry Perspective

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

Uday Kumar · Alireza Ahmadi
Ajit Kumar Verma · Prabhakar Varde
Editors

Current Trends in Reliability, Availability, Maintainability and Safety

An Industry Perspective

Springer

Editors

Uday Kumar
Operation and Maintenance Engineering
Luleå University of Technology
Luleå
Sweden

Alireza Ahmadi
Operation and Maintenance Engineering
Luleå University of Technology
Luleå
Sweden

Ajit Kumar Verma
ATØM, University College
Haugesund
Norway

Prabhakar Varde
Research Reactor Services Division
Bhabha Atomic Research Centre
Mumbai
India

ISSN 2195-4356 ISSN 2195-4364 (electronic)
Lecture Notes in Mechanical Engineering
ISBN 978-3-319-23596-7 ISBN 978-3-319-23597-4 (eBook)
DOI 10.1007/978-3-319-23597-4

Library of Congress Control Number: 2015954616

Springer Cham Heidelberg New York Dordrecht London

© Springer International Publishing Switzerland 2016
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

Preface

The ICRESH-ARMS 2015 Conference at Luleå was an important milestone in the history of activities of the Division of Operation and Maintenance, Luleå Technical University. It was hosted by the Operation and Maintenance Division at LTU and had considerable participation both from the division and the university. This conference was also successful in attracting a fairly large number of researchers worldwide to deliver invited talks, present papers, and attend the exciting sessions. There was a good interest in participation from the industries around that saw an opportunity to interact with engineers, researchers, academicians, and managers who are at the forefront of related technologies in reliability, operation and maintenance, condition monitoring, risk and safety in various domains, eg. Manufacturing, Transportation, Defense, Power, Mining, and the IT sector. In particular, there was significant interest in Railway infrastructure and asset management. Condition-based maintenance with emphasis on diagnostics, prognostics and health management, and maintenance of large engineering systems and their implications to risk and safety were another area where the industry and academia found common interest. A growing interest in the use of nonrenewable energy sources saw several presentations from India highlighting various risk and safety issues besides quantitative risk assessment. Some papers reflect active interest in industry in big data analytics and context-based thinking for decision making and signified a growing awareness amongst industries to adapt to new technologies. Mining and Railways were two domains with significant local interest in Sweden and saw various presentations in innovative approaches to long-term solutions related to aging, health management, and prognostics. Modeling and Simulation saw increasing applications in various industrial domains and a growing trend among researchers to develop a holistic and integrated approach to address various conflicting issues of reliability, risk, production, and cost. An important highlight was the discussions that the interactive sessions triggered during the breaks between the participants from the industry and the academicians. There were in all 82 papers selected after reviews for presentation in various sessions, namely:

- Degradation/Aging and Preventive Maintenance
- Diagnostics, Prognostics, and Health Management
- Maintenance Management
- Maintenance Modeling and Analysis
- Performance Management and Energy
- Software Reliability
- Probabilistic Risk and Safety Analysis
- Reliability Analysis and Modeling
- Reliability and Maintenance of Mining Machinery

A total of 53 papers have been selected in this book from across all the sessions in the conference. They cover a large spectrum of the theme of the conference and present exciting findings which enrich the current applied knowledge base in the subject with indicators to the future in the industry. The endeavor in the selected compilation has been to avoid losing focus of the ultimate beneficiary of the applied nature of this work in the industry in various domains.

On behalf of the ICRESH-ARMS organization, we wish to express our sincere appreciation to the conference delegates, the distinguished keynote speakers, authors, workshop leaders, and members of the scientific review committee of the ICRESH-ARMS 2015, for their outstanding contribution towards the success of the conference. We are thankful also to Dr. Adithya Thaduri, for his support to prepare the proceedings.

Editorial Board

Dr. Uday Kumar
Dr. Alireza Ahmadi
Dr. Ajit Kumar Verma
Dr. Prabhakar Varde

Organization of the Conference

Patron/Honorary chair

Professor Johan Sterte, rektor, Luleå University of Technology, Sweden

Conference General Chairs

Uday Kumar, Luleå University of Technology, Sweden
R.C. Sharma, BARC, Mumbai-India

Conference Chairs and Conveners

P-O Larsson Kråik, Trafikverket-Sweden
Alireza Ahmadi, Luleå University of Technology, Sweden

International Organizing Chairs

P.V. Varde, BARC-India
A.K. Verma, HSH-Norway
Diego Galar, Luleå University of Technology, Sweden

Scientific Chairs

A. Srividya, HSH-Norway
Behzad Ghodrati, Luleå University of Technology, Sweden

Industry Relation Chairs

Ramin Karim, Luleå University of Technology, Sweden
Olov Candell, SAAB-Sweden

Workshop and Tutorial Chairs

Aditya Parida, Luleå University of Technology, Sweden
Diganta Das, Calce, UMD-USA

Program and Public Relation Chairs

Philip Tretten, Luleå University of Technology, Sweden
Veronica Jagare, JVTC-Sweden

Secretaries of Administration and Finance

Alexandra Lund Cipolla, Luleå University of Technology, Sweden
Katarina Grankvist, Luleå University of Technology, Sweden
Cecilia Glover, Luleå University of Technology, Sweden

Editorial Committee

Uday Kumar, Luleå University of Technology, Sweden
Alireza Ahmadi, Luleå University of Technology, Sweden
A.K Verma, HSH-Norway
P.V. Varde, BARC-India

Executive Committee

Matti Rantatalo, LTU-Sweden
Iman Arasteh Khouy, LTU-Sweden
Yamur Aldouri, LTU-Sweden
Adithya Thaduri, LTU-Sweden
Christer Stenström, LTU-Sweden
Stephen Mayowa Famurewa, LTU-Sweden
Amparo Morant, LTU- Sweden
Hussan Hamoudi, LTU-Sweden
Mustafa Aljumaili, LTU-Sweden
Hadi Hosseini, LTU-Sweden
Amir Soleimani, LTU- Sweden

Scientific Committee

Mohammed Al Yahyai, UK
John Arul, India
Morteza Bagheri, Iran
J.C. Bansal, India
Javad Barabady, Norway
Jorge Baron, Argentina
Mohamed Ben-Daya, USA
Neil Blundell, France
Luciano Burgazzi, Italy
Gopinath Chattopadhyay, Australia
Mohammad Ali Farsi, Iran
Veronica Garea, Argentina
Pieter Gelder, Belgium
V. Gopika, India
Suprakash Gupta, India
Achintya Haldar, USA
Isabelle Hendrickx, Belgium
Philippe Hessel, Canada
Mattiass Holmgren, Sweden
Kouroush Jenab, USA
P.K. Kapur, India
Rezaul Karim, Bangladesh
Kevin Koyan, USA
Manoj Kumar, India
Senthil Kumar, India
Jeanne-Marie Lanore, France

Janet Lin, Sweden
J.P. Liyanage, Norway
Ramón M. López, Mexico
Jan Lundberg, Sweden
Tore Markeset, Norway
Joseph Mathew, Australia
Mohammad Modaress, USA
Mohammad Mosleh, USA
Pra Murthy, Australia
V.N.A. Naikan, India
Arne-Nisen, Sweden
Johan Odelius, Sweden
S. Osaki, Japan
Mahesh Pande, Canada
Ljubisa Papic, Serbia
Michael Pecht, USA
Mohammad Pourgol-Mohammad, Iran
Raghu Prakash, India
Matti Rantatalo, Sweden
Mohsen Rezaeian, Iran
Marina Röwekamp, Germany
Håkan Schunnesson, Sweden
Rehan Sediq, Canada
Oliver Straeter, Germany
Peter Söderholm, Sweden
Bernardo Tormos, Spain
Ali Turkyilmaz, Turkey
Davood Younesian, Iran
Sufian Yousef, UK
Jabbar Ali Zakeri, Iran

Contents

Part I Degradation/Aging & Preventive Maintenance

A Survey on Track Geometry Degradation Modelling 3
Iman Soleimanmeigouni and Alireza Ahmadi

Maintenance Optimization Using Multi-attribute Utility Theory 13
A.H.S. Garmabaki, Alireza Ahmadi and Mahdieh Ahmadi

Optimum Proactive Maintenance for Critical Infrastructures Subject to Multiple Degradation and Environmental Shocks 27
Mahmood Shafiee

Risk Informed In-Service Inspection of PWR Nuclear Power Plant Piping Components Subjected to Erosion-Corrosion Using Markov Chain Model 39
K. Balaji Rao, M.B. Anoop, Gopika Vinod and H.S. Kushwaha

Turnout Degradation Modelling Using New Inspection Technologies: A Literature Review 49
Niloofar Minbashi, Morteza Bagheri, Amir Golroo, Iman Arasteh Khouy and Alireza Ahmadi

Part II Diagnostics, Prognostics and Health Management

Context-Based Maintenance and Repair Shop Suggestion for a Moving Vehicle 67
Adithya Thaduri, Diego Galar, Uday Kumar and Ajit Kumar Verma

Optimal Sensor Placement for Efficient Fault Diagnosis in Condition Monitoring Process; A Case Study on Steam Turbine Monitoring 83
Farzin Salehpour Oskouei and Mohammad Pourgol-Mohammad

Estimation of the Reliability of Rolling Element Bearings Using a Synthetic Failure Rate 99
Urko Leturiondo, Oscar Salgado and Diego Galar

Nonlinear Process Monitoring Using Genetic Algorithms 113
Tawfik Najeh, Achraf Jaber Telmoudi and Lotfi Nabli

Vibration and Acoustics Emissions Analysis of Helicopter Gearbox, A Comprative Study 125
Faris Elasha and David Mba

Test Rig Assessment of an On-Line Wear Sensor for Application in Wind Turbine Gearboxes 141
Vicente Macián, Bernardo Tormos, Santiago Ruiz, Guillermo Miró and Isaac Rodes

Residual Signal Techniques Used for Gear Fault Detection 157
Omar D. Mohammed and Matti Rantatalo

Prognostics and Structural Health Assessment Using Uncertain Measured Response Information 165
Achintya Haldar and Abdullah Al-Hussein

Ontology Based Diagnosis for Maintenance Decisions of Paper Mill Roller Using Dynamic Response. 187
Madhav Mishra and Adithya Thaduri

Context Awareness in Predictive Maintenance 197
Bernard Schmidt, Diego Galar and Lihui Wang

Prognostics and Health Management: Methodologies & Soft Computing Techniques 213
S.V. Shrikhande, P.V. Varde and D. Datta

Intelligent Real-Time Risk Analysis for Machines and Process Devices 229
Esko K. Juuso and Diego Galar

Part III Maintenance Management

Malfunction in Railway System and Its Effect on Arrival Delay 243
Sida Jiang and Christer Persson

On-Condition Parts Versus Life Limited Parts: A Trade off in Aircraft Engines. 253
Veronica Fornlöf, Diego Galar, Anna Syberfeldt and Torgny Almgren

In Need for Better Maintenance Cost Modelling to Support the Partnership with Manufacturing 263
Gary Linnéusson, Diego Galar and Mikael Wickelgren

Investigation of Causes of Mining Machines Maintenance Problems . . . 283
Ljubisa Papic, Srdja Kovacevic, Diego Galar and Adithya Thaduri

Part IV Maintenance Modeling and Analysis

Safety and Availability Evaluation of Railway Signalling Systems 303
Amparo Morant, Anna Gustafson and Peter Söderholm

Applying Multi-factorial Pareto Analysis in Prioritizing Maintenance Improvement Initiatives. 317
Marcus Bengtsson

Predictive Modelling for Estimation of Railway Track Degradation. . . . 331
Bjarne Bergquist and Peter Söderholm

Facilitating the Maintenance of Safety Cases 349
Omar Jaradat, Iain Bate and Sasikumar Punnekkat

Track Maintenance Between Trains Simulation 373
Christer Stenström, Ulla Juntti, Aditya Parida and Uday Kumar

Process Analysis of Human Failures in Railway Maintenance 381
Mattias Holmgren and Peter Söderholm

Application of Game Theory to Railway Decision Making 395
N. Jack, D.N.P. Murthy and U. Kumar

Modelling of Maintenance Data. 409
M.R. Karim, A. Ahmadi and D.N.P. Murthy

Scheduling of Railway Infrastructure Maintenance Tasks Using Train Free Windows 425
Stephen M. Famurewa, Arne Nissen and Uday Kumar

A Survey on Predictive Maintenance Through Big Data 437
Amit Patwardhan, Ajit Kumar Verma and Uday Kumar

Part V Probabilistic Risk and Safety Analysis

Approach for Probabilistic Safety Assessment of Accelerator 449
Gopika Vinod, M. Hari Prasad, G. Haridas and R.K. Singh

A Comparative Risk Assessment for Sites with Single and Double Units 459
Varun Hassija, C. Senthil Kumar, K. Velusamy and V. Balasubramaniyan

Numerical Analysis of a Railway Compartment Fire. 471
Anwar Enbaya, Taimoor Asim, Rakesh Mishra
and Raj B.K.N. Rao

Safety Analysis of Mining Machines Specific Maintenance Operations 485
Ljubisa Papic, Srdja Kovacevic, Diego Galar and Adithya Thaduri

Experiences and Insights in Development of Probabilistic Safety Assessment of Research Reactors in BARC 497
Sachin Kumar, N.S. Joshi, Vivek Mishra and P.V. Varde

Part VI Reliability Analysis and Modeling

Reliability Model of a Safety System with an Imperfect Tester 511
Pramod Kumar Sharma and A. John Arul

Rail Breaks—An Explorative Case Study 519
Peter Söderholm and Bjarne Bergquist

An Integrated Approach to Remaining Life Prediction of Rotating Machines 543
Tarun Chugh, V. Sankaranarayanan and P.V. Varde

Reliability Compliance Testing of Electronic Systems Using Parametric and Non Parametric Sequential Test Plans 561
Diana Denice, Manoj Kumar and P.P. Marathe

Bayesian Reliability with MCMC: Opportunities and Challenges 575
Jing Lin

Root Cause Analysis in Support of Event Investigation 587
N.S. Joshi, Sachin Kumar and P.V. Varde

Maintenance of Large Engineering Systems 599
Anil Rana, Ajit Kumar Verma and Ajit Srividya

Maintenance Risk Based Inspection Optimization Model in Multi-Component Repaiable System with Economic Failure Interaction 611
Esmaeil Rezaei and Din Mohammad Imani

Part VII Reliability and Maintenance of Mining Machinery

Reliability Analysis and Maintenance Scheduling of the Electrical System of Rotary Drilling Machines 623
Mohammad Javad Rahimdel, Mohammad Ataei and Reza Khalokakaei

Monte Carlo Reliability Simulation of Underground Mining Drilling Rig 633
Hussan Al-Chalabi, Hadi Hoseinie and Jan Lundberg

Comparison of Mine Production Index Factors for Rock Bolter and Shovel 645
Amol Lanke and Behzad Ghodrati

Reliability Analysis of Face-to-Surface Continuous Coal Hauling System in Longwall Mines 657
Amid Morshedlou and Hesam Dehghani

Simulation of an Active Maintenance Policy: A Preliminary Study in Dragline Maintenance Optimization 669
Onur Gölbaşı and Nuray Demirel

Reliability Analysis of Motor System of Dump Truck for Maintenance Management 681
Zeynab Allahkarami, Ahmad Reza Sayadi and Amol Lanke

Part VIII Software Reliability & Data Quality

Multi Up-Gradation Reliability Model for Open Source Software 691
Mahdieh Ahmadi, Iraj Mahdavi and A.H.S. Garmabaki

Quality of Streaming Data in Condition Monitoring Using ISO 8000 703
Mustafa Aljumaili, Ramin Karim and Phillip Tretten

Data Quality Assessment of Automatic Wheel Profile Measurement Systems 717
Matthias Asplund, Stephen M. Famurewa and Matti Rantatalo

Part I
Degradation/Aging & Preventive Maintenance

A Survey on Track Geometry Degradation Modelling

Iman Soleimanmeigouni and Alireza Ahmadi

Abstract Railway transportation is exposed to a higher demand that necessitates the use of trains with higher speed and heavier axle loads. These increase the track geometry degradation rate, which needs a more effective control on geometry degradation. Keeping the track geometry in acceptable levels requires proper inspection and maintenance planning that inevitably entails in-depth knowledge of track geometry degradation. In addition, it is needed to identify the most effective approach for degradation modelling. To do so, it is vital to synthesis published results into a summary of what is known and validated and what is not as a major step. To this end, this paper reviews track degradation models, discusses various degradation measures, and proposes directions for future researches. It is found that combining the mechanistic and statistical approaches can leads to a more accurate prediction of track geometry degradation behaviour.

Keywords Railway track · Maintenance modelling · Degradation model · Degradation measures

1 Introduction

The railway track and infrastructure degrade with age and usage and can become unreliable due to failure. When a failure occurs, the consequences can be significant, including a high cost of railway operation, economic loss, damage to the railway asset and environment and possible loss of human lives. Unreliability may also lead to annoyance, inconvenience and a lasting customer dissatisfaction that

I. Soleimanmeigouni (✉) · A. Ahmadi
Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden
e-mail: iman.meigouni@ltu.se

A. Ahmadi
e-mail: alireza.ahmadi@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_1

can create serious problems for the company's position in the marketplace. An applicable and effective maintenance strategy can guarantee the achievement of reliability goals and compensate for unreliability.

Maintenance actions are used to control the degradation of the track, reduce or eliminate the likelihood of failures, and restore a failed part to an operational state. We need to model track degradation behaviour if we are to select an applicable and effective maintenance policy, but modelling and predicting the track geometry degradation is a complex task, requiring the following information: (1) the interaction of different track components, (2) the effect of maintenance actions on track quality, (3) the heterogeneity factors e.g. environmental factors, soil type and condition.

In addition, higher demand for railway transportation creates an essential requirement for higher speed and axle load which accelerates the track aging process and negatively affects its reliability.

The increased demand and complexity dictates the need for comprehensive track degradation models. More studies are needed to provide a scientific footing for track degradation modelling. Synthesising the published results into a summary of what is known and validated and what is not is a major step. Therefore, the aim of this paper is to review the literature on track degradation modelling. The paper offer insights into the different construction of track geometry degradation problems and modelling to formulate questions that need further research.

The reminder of the paper is organised as follow. Section 2 describes the track structure. Section 3 discusses recent works on track geometry degradation modelling. The discussion and conclusion are provided in Sect. 4.

2 Track Structure

The majority of railway tracks around the world are ballasted tracks, which is the interest of this study. The conventional ballasted tracks have lower construction cost, and properly respond to different static and dynamic force [1]. The ballasted-track components are demonstrated in Fig. 1. The static and dynamic forces transform through wheels to rail and consequently to sleepers, ballast, sub-ballast, and finally sub-grade. Rails are longitudinal steel members that guide train wheels and distribute the forces from train wheels to sleepers. Fasteners are used to fix rails to sleepers and prevent longitudinal, vertical, and lateral movements of rails [1]. Sleepers provide a solid and flat support for rails and keep it within acceptable positions along the track using fasteners. In addition, sleepers distribute the vertical, longitudinal, and lateral forces due to rail-wheel contact to the ballast [1, 2]. Three sleeper types can be found around the world, i.e. steel, wooden, and concrete sleepers. A comprehensive study about sleeper types and their failure modes is conducted by Ferdous et al. [3]. Ballast is consisted of crushed stones and its overall goal is to transmit the forces from the sleepers to sub-ballast layer. Ballast prevents the track to exit from acceptable lateral, vertical, and longitudinal

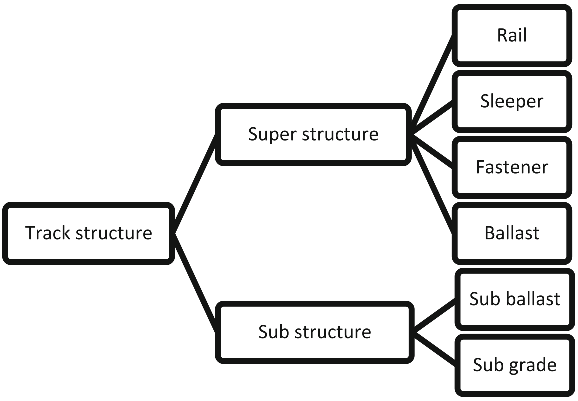


Fig. 1 Track components

thresholds [1, 4]. Sub ballast is a layer to transmit the forces from super structure to sub grade. Sub ballast decreases the stresses at the end of ballast to avoid probable damage to sub grade surface. In addition, sub ballast prevent from entrance of sub grade materials into ballast layer that reduce drainage efficiency of ballast [1, 5].

Subgrade is a surface of ground that uses as foundation to construct track structure. In some cases the ground can be modified with special materials to remove profile problems. Sub grade plays a key role in supporting track system and a failure in sub grade will generate massive consequences, even with high quality ballast and sub ballast [1, 5].

3 Survey of Track Geometry Degradation Models

In the two last decades, a great deal of research has been done in the field of track geometry degradation modelling. Determining an indicator to represent track quality is an essential prerequisite for modelling track degradation. Different indicators are used based on the aim of the research. The indices for representing track quality condition are demonstrated in Fig. 2.

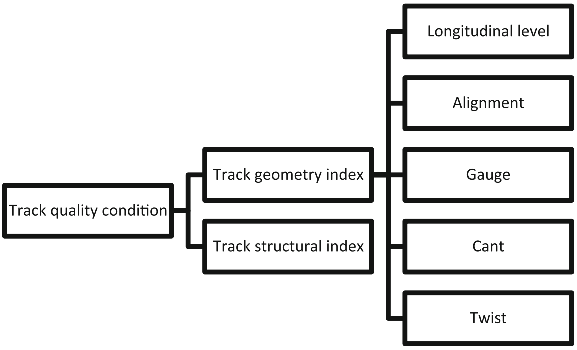


Fig. 2 Track condition measures

Sadeghi et al. [6] proposed a track geometry index uses the following track geometry parameters: alignment, profile, twist, gauge, and rail cant. Using justified coefficient, they combined the parameters to design the track geometry index. Sadeghi [7] developed the previous work by proposing an overall track geometry index by considering different track classes.

In order to consider structural defects, Sadeghi et al. [8] proposed a quantitative track structural quality index. This index is defined for each track component group, i.e. rail, sleeper, fastener, ballast.

Later, Sadeghi et al. [9] used the neural network technique to correlate track geometry irregularities to track structural defects.

Berawi et al. [10] compared three track quality condition measures: J Synthetic, Indian TGI, and a measure based on European Standard EN 13848-5. They observed that different track evaluation methods resulted in different degradation rates, with the TGI having the highest degradation rate of the three. Faiz et al. [11] studied the geometry parameters used in the UK track maintenance process and applied linear regression analysis to explain their correlations. A Generalized Energy Index (GEI) instead of a Track Quality Index (TQI) for track quality evaluation is proposed by Li et al. [12]. The GEI can consider different track irregularity wave-length and speed. Haifeng et al. [13] proposed an integral maintenance index (IMI) that considers the distribution of track geometry parameters to evaluate track condition. El-Sibaie et al. [14] developed a number of track quality indices to evaluate track quality condition in relation to different track classes.

By looking to the literature it can be observe that most of the researchers considered short wavelength longitudinal level as the crucial factor in degradation modelling. This issue can be seen in Fig. 3.

After finding the proper track quality measure, a degradation model must be constructed and the effect of different maintenance strategies on track degradation evaluated. There are two major approaches for track geometry degradation

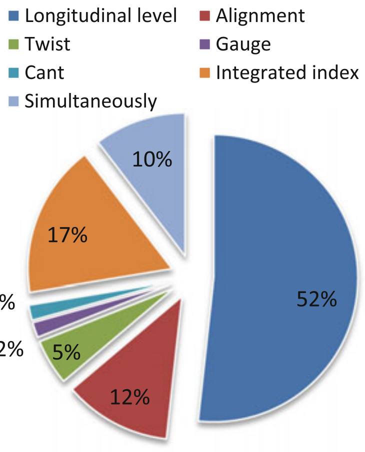


Fig. 3 Distribution of applied track geometry measures

modelling, i.e. mechanistic and statistical approaches. In this paper, statistical approach is the main focus.

Concerning mechanistic approach, a number of researchers tried to find the interactions among track components and their influences on track geometry degradation.

The most important models are those proposed by Shenton [15], Sato [16, 17], Chrismer et al. [18], Öberg et al. [19], and Zhang et al. [20]. Dahlberg [21] also provide an extensive review on mechanistic models applied for track geometry degradation.

Concerning statistical approaches, the most commonly applied methods are summarised in Fig. 4.

A stochastic Markov model is used by Bai et al. [22] to evaluate track degradation. They considered various heterogeneous factors and argued that the existence of these factors caused two maintenance units with the same mileage to show different degradation behaviour. A Markov model is deployed by Yousefikia et al. [23] to model tram track degradation and obtain the optimal maintenance strategy. A model by integrating the grey model and Markov chain is developed by Liu et al. [24] to predict track quality condition. Xu et al. [25] proposed a track measures data mining model to predict railway track degradation for a short time period. A framework called the tree-augmented naïve Bayes-track quality index is proposed by Bai et al. [26] to predict railway track irregularities for short-term horizon.

Guler [27] used artificial neural networks to model the degradation of different track geometry parameters. The model considered traffic load, velocity, curvature, gradient, cross-level, sleeper type, rail type, rail length, falling rock, land slide, snow, and flood as influencing factors. A multi-stage linear model is applied by Gou et al. [28] to cope with different phases of degradation between two consecutive maintenance interventions and the exponential growth of track irregularity.

Famurewa et al. [29] compared the accuracy of linear, exponential, and grey models in the estimation and prediction of track geometry degradation. The comparison demonstrated the grey model has lower mean average percentage error than the linear model and an approximately equal error value with the exponential model. The Gaussian random process is used by Zhu et al. [30] to model track irregularities in vertical profile and alignment. They discussed power spectral density analysis and cross-level statistics about track irregularities to improve track degradation modelling.

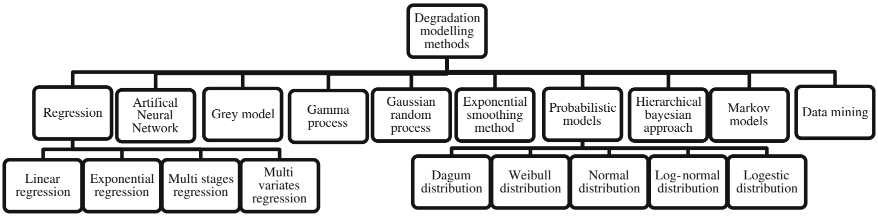


Fig. 4 Track degradation approaches

Using waveform data, Liu et al. [31] proposed a short range prediction model to estimate any track irregularity index over a short track section length (25 m) and on a day-by-day basis. They concluded the total process of track surface change over track sections is nonlinear and different track sections have different nonlinear process.

Data mining and time series theories are applied by Chaolong et al. [32] to predict track irregularity standard deviation time series data. In order to predict the changing trends of track irregularity, they used the linear recursive model and the linear autoregressive moving average model.

A modified grey model is developed by Chaolong et al. [33] to analyse track irregularity time series data and obtain a medium-long term prediction of track cross levelling. They compared the stochastic linear autoregressive model, Kalman filtering model, and artificial neural network with respect to the short term track cross levelling prediction. They observed the accuracy of the ANN model was higher than the two other models. A stochastic approach based on Dagum distribution is developed by Vale et al. [34] to model track longitudinal level degradation over time. The researchers classified the track longitudinal level changes into three speed classes and different inspection intervals.

Andrade et al. [35] used a Bayesian approach to evaluate a track geometry degradation model and deal with the uncertainty of its parameters. They considered the track longitudinal level deviation to have a linear relationship with passing tonnage and assumed the initial longitudinal level and degradation rate would take a bivariate log-normal prior distribution. They argued that the parameter uncertainties are significant in the design stage.

In order to model track geometry degradation and maintenance, Westgeest et al. [36] addressed the application of regression method. They used a combination of track geometry parameters to create the Key Performance Indicator (KPI) as the track quality indicator. They studied the effect on the KPI of different types of subsoil, sleeper, tonnage, and engineering structures, considering two tamping types, manual and mechanical. The results showed the proposed degradation model can properly address changes in the KPI over time, but it is not efficient in terms of track behaviour prediction. They concluded the track segments have different degradation rates depending on a number of factors, e.g. closeness to switches, sleeper types, and subsoil types.

Xu et al. [37] proposed an approach based on historical changes in track irregularity to predict the short-term track degradation. They estimated the non-linear behaviour of track irregularity during a cycle using a number of short range linear regression models. Andrade et al. [38] assessed track geometry degradation and the uncertainty of degradation model parameters. They considered a linear model for track longitudinal level degradation. They performed statistical correlation analysis for each group section and fitted the log-normal distribution to the track's longitudinal level degradation.

A machine learning model based on the characteristics and inspection data of the track using a multi-stage framework is developed by Xu et al. [39] to predict changes in track irregularity over time. They defined different stages of track irregularity changes based on maintenance thresholds and linear regression is used to predict track degradation in each stage.

Berggren [40] applied the pattern recognition method to extract new information from track quality condition data to classify the root cause of track problems. He defined three classes of defects: rail defects, ballast defects, and soil defects. The proposed framework also used data related to track geometry quality, dynamic stiffness, and ground penetrating radar. The main output of the framework is classifying the feature measurements based on their effect on track problems.

The application of multivariate statistical analysis for geometry degradation modelling is pointed out by Guler et al. [41]. First, they divided the track into homogenous sections based on gradient, curvature, cant, speed, age, rail type, and rail length. They examined the effect of traffic load, speed, curvature, gradient, cant, sleeper type, rail type, rail length, falling rock, land-slide, snow, and flood. They concluded landslide and snow do not affect track geometry degradation, but rail type and rail length do. The model found a high correlation between cant and curvature.

Chang et al. [42] proposed a multi-stage linear model to predict changes in track irregularity. Based on multi-stage and exponential changes in track irregularities, they modelled different stages of TQI changes using a number of linear models. The different stages of track irregularity changes were based on the TQI distribution.

The comparison of the efficiency of the double exponential smoothing method, a generic degradation model, and an autoregressive model for track degradation prediction is addressed in the work by Quiroga et al. [43]. The three models lose their efficiency in track degradation prediction after performing a number of tamping procedures. After considering these issues, they developed a hybrid discrete-continuous framework based on a grey box model. After comparing these four models, they concluded the proposed hybrid model is more efficient in terms of track degradation behaviour prediction.

A degradation model by combining mechanistic and statistical approaches based on regression that considered track geometry and track structural condition data is proposed in the work by Sadeghi et al. [44]. Using a degradation coefficient, they estimated the effect of initial track geometry and track structural condition, train speed, and total million gross tones passing on the track. They observed an exponential relationship between the degradation coefficient and the parameters. The initial track quality condition was found to be the most effective parameter acting on the degradation coefficient, with the total million gross tonnes passing on the track coming second. They concluded the degradation coefficient is more affected by parameters in turnouts, bridges, and curve-bridges than by parameters in other track segment types.

Lyngby [45] suggested a methodology for evaluating track degradation in terms of track geometry irregularities and proposed a multivariate regression model to demonstrate the relationship between the track degradation measure variable and influencing variables on track degradation. Since different sections of track are not identical, the track was split into homogenous sections with similar variables. He concluded: (1) axle load has a nonlinear relation with degradation; (2) degradation after tamping is dependent on the number of previous tampings; (3) soil consisting of clay material will settle sooner than other types of soil; (4) light rail tracks degrade faster than heavy rail tracks; (5) harsh rainfall increases degradation rate.

Two degradation models to predict track alignment irregularities are proposed in the work by Kawaguchi et al. [46]. First, they developed a degradation model based on analysis of lateral track deformation to estimate mean time to maintenance of track alignment irregularities. Second, they designed another degradation model based on the exponential smoothing method to accurately predict the track alignment irregularities a maximum of 1 year in advance.

A generic degradation model is developed by Jovanovic [47], which is suitable for modelling degradation of different parameters. To develop a generic degradation model, he argued the condition parameters that represent the condition of track components and essential and temporary activities affecting them should be determined. He observed different degradation patterns based on various intervals between essential or temporary activities. Various curve types, such as linear, exponential, and quadratic, can be used to explain the degradation patterns. Miwa et al. [48] fitted the logistic distribution on track irregularity data to express the track condition, with the parameters of the distribution related to type of alignment, rail, and sleeper; depth of ballast; and maintenance history using the exponential smoothing method.

4 Conclusion

The paper aims to improve the knowledge of railway track geometry degradation modelling by conducting a survey of recent research works. The most important issues to consider in track modelling are identified.

Since sections are not homogenous along the track and different heterogeneous factors affect track degradation, the track must be divided into a number of shorter track sections and maintenance must be planned for each section. It is observed that constant equal length sections are usually considered for planning track maintenance activities. However, a more efficient method would be clustering sections based on their structural, environmental, and operational characteristics.

In addition, achieving a more accurate prediction of track geometry degradation requires combining the mechanistic and statistical approaches. Finally, integrating the degradation models of different track components, i.e. ballast, rail, and sleeper, to plan track maintenance activities can increase the efficiency of maintenance strategies in terms of economy of scale.

After modelling the track geometry degradation, the track geometry maintenance models could be constructed. In fact, by combining the degradation and restoration models the long term behaviour of the track could be predicted. In this regard, infrastructure managers can evaluate different maintenance strategies.

References

1. Lyngby N, Hokstad P, Vatn J (2008) RAMS management of railway tracks. In: Handbook of performability engineering. Anonymous: Springer, pp 1123–1145
2. Li G, Huang Z (2008) A study on reliability of railway sleeper system based-on finite Markov chain imbedding approach. In: 4th international conference on wireless communications, networking and mobile computing. WiCOM'08, pp 1–4
3. Ferdous W, Manalo A (2014) Failures of mainline railway sleepers and suggested remedies–review of current practice. Eng Fail Anal 44:17–35
4. Prescott D, Andrews J (2013) Modelling maintenance in railway infrastructure management. In: 2013 Proceedings-annual reliability and maintainability symposium (RAMS), pp 1–6
5. Tzanakakis K (2013) Earliest Traces. In: The Railway Track and Its Long Term Behaviour. Springer, Anonymous, pp 3–6
6. Sadeghi J, Fathali M, Boloukian N (2009) Development of a new track geometry assessment technique incorporating rail cant factor. Proc Inst Mech Eng Part F J Rail Rapid Transit 223 (3):255–263
7. Sadeghi J (2010) Development of railway track geometry indexes based on statistical distribution of geometry data. J Transp Eng 136(8):693–700
8. Sadeghi J, Askarinejad H (2011) Development of track condition assessment model based on visual inspection. Struct Infrastruct Eng 7(12):895–905
9. Sadeghi J, Askarinejad H (2012) Application of neural networks in evaluation of railway track quality condition. J Mech Sci Technol 26(1):113–122
10. Berawi ARB, Delgado R, Calçada R, Vale C (2010) Evaluating track geometrical quality through different methodologies. Int J Technol 1(1):38–47
11. Faiz R, Singh S (2009) Time based predictive maintenance management of UK rail track. In: ICC'09. International Conference on computing, engineering and information, pp 376–383
12. Li H, Xiao T (2014) Improved generalized energy index method for comprehensive evaluation and prediction of track irregularity. J Stat Comput Simul 84(6):1213–1231
13. Li H, Xu Y (2009) Railway Track Integral Maintenance Index and Its Application. Int Conf Transp Eng 2009:2514–2519
14. El-Sibaie M, Zhang Y (2004) Objective track quality indices. Transp Res Record J Transp Res Board 1863:81–87
15. Shenton M (1985) Ballast deformation and track deterioration. Track Technol 253–265
16. Sato Y (1997) Optimum track structure considering deterioration in ballasted track. In: Proceedings of 6th international heavy haul conference, Cape Town, SouthAfrica, p 576
17. Sato Y (1995) Japanese studies on deterioration of ballasted track. Veh Syst Dyn 24 (sup1):197–208
18. Chrismer S, Selig E (1993) Computer model for ballast maintenance planning. In: Proceedings of 5th International Heavy Haul Railway Conference, Beijing, China, pp 223–227
19. Öberg J, Andersson E (2009) "Determining the deterioration cost for railway tracks. Proc Inst Mech Eng Part F J Rail Rapid Transit 223(2):121–129
20. Zhang Y, Murray MH, Ferreira L (2000) Modelling rail track performance: an integrated approach. Transp J 187–194
21. Dahlberg T (2001) Some railroad settlement models—a critical review. Proc Inst Mech Eng Part F J Rail Rapid Transit 215(4):289–300
22. Bai L, Liu R, Sun Q, Wang F, Xu P (2015) "Markov-based model for the prediction of railway track irregularities. Proc Inst Mech Eng Part F J Rail Rapid Transit 229(2):150–159
23. Yousefikia M, Moridpour S, Setunge S, Mazloumi E (2014) Modeling degradation of tracks for maintenance planning on a tram line. J Traffic Logist Eng 2(2)
24. Liu F, Wei X, Liu Y, Yin X (2014) Track quality prediction based on center approach Markov-Grey GM (1, 1) model. In: 2014 IEEE international conference on information and automation (ICIA), pp 81–86

25. Xu P, Liu R, Wang F, Wang F, Sun Q (2013) Railroad track deterioration characteristics based track measurement data mining. Math Probl Eng 2013
26. Bai L, Liu R, Sun Q, Wang F, Wang F (2014) Classification-learning-based framework for predicting railway track irregularities. Proc Inst Mech Eng Part F J Rail Rapid Transit 0954409714552818
27. Guler H (2014) Prediction of railway track geometry deterioration using artificial neural networks: A case study for turkish state railways. Struct Infrastruct Eng 10(5):614–626
28. Guo R, Han BM (2013) Multi-stage linear prediction model for railway track irregularity. Appl Mech Mater 361:1811–1816
29. Famurewa SM, Xin T, Rantatalo M, Kumar U Comparative study of track geometry quality prediction models
30. Zhu M, Cheng X, Miao L, Sun X, Wang S (2013) Advanced stochastic modeling of railway track irregularities. Adv Mech Eng 5:401637
31. Liu R, Xu P, Wang F (2010) Research on a short-range prediction model for track irregularity over small track lengths. J Transp Eng 136(12):1085–1091
32. Chaolong J, Weixiang X, Lili W, Hanning W (2013) Study of railway track irregularity standard deviation time series based on data mining and linear model. Math Probl Eng 2013
33. Chaolong J, Weixiang X, Futian W, Hanning W (2012) Track irregularity time series analysis and trend forecasting. Discret Dyn Nat Soc 2012
34. Vale C, Lurdes SM (2013) Stochastic model for the geometrical rail track degradation process in the portuguese railway northern line. Reliab Eng Syst Saf 116:91–98
35. Andrade AR, Teixeira PF (2012) A bayesian model to assess rail track geometry degradation through its life-cycle. Res Transp Econ 36(1):1–8
36. Westgeest F, Dekker R, Fischer R (2011) Predicting rail geometry deterioration by regression models. Adv Saf Reliab Risk Manag ESREL 2011:146
37. Xu P, Sun Q, Liu R, Wang F (2011) A short-range prediction model for track quality index. Proc Inst Mech Eng Part F J Rail Rapid Transit 225(3):277–285
38. Andrade AR, Teixeira PF (2011) Uncertainty in rail-track geometry degradation: Lisbon-oporto line case study. J Transp Eng
39. Xu P, Liu R, Wang F, Sun Q, Teng H (2011) A novel description method for track irregularity evolution. Int J Comput Intell Syst 4(6):1358–1366
40. Berggren E (2010) Efficient track maintenance: methodology for combined analysis of condition data. Proc Inst Mech Eng Part F J Rail Rapid Transit 224(5):353–360
41. Guler H, Jovanovic S, Evren G (2011) Modelling railway track geometry deterioration. Procee ICE Transp 164(2):65–75
42. Chang H, Liu R, Wang W (2010) Multistage linear prediction model of track quality index. In: Proceeding the conference on traffic and transportation studies. ICTTS, Kunming, pp 1183–1192
43. Quiroga L, Schnieder E (2010) Modelling high speed railroad geometry ageing as a discrete-continuous process. In: Proceedings of the stochastic modeling techniques and data analysis international conference, SMTDA
44. Sadeghi J, Askarinejad H (2010) Development of improved railway track degradation models. Struct Infrastruct Eng 6(6):675–688
45. Lyngby N (2009) Railway track degradation: Shape and influencing factors. Int J Perform Eng 5(2):177
46. Kawaguchi A, Miwa M, Terada K (2005) Actual data analysis of alignment irregularity growth and its prediction model. Q Rep RTRI 46(4):262–268
47. Jovanovic S (2004) Railway track quality assessment and related decision making. In: 2004 IEEE international conference on systems, man and cybernetics, pp 5038–5043
48. Miwa M, Ishikawa T, Oyama T (2000) Modeling the transition process of railway track irregularity and its application to the optimal decision-making for multiple tie tamper operations. In: Proceedings of the international conference, railway engineering 2000, held London, UK, July 2000-CD-ROM

Maintenance Optimization Using Multi-attribute Utility Theory

A.H.S. Garmabaki, Alireza Ahmadi and Mahdieh Ahmadi

Abstract Several factors such as reliability, availability, and cost may consider in the maintenance modeling. In order to develop an optimal inspection program, it is necessary to consider the simultaneous effect of above factor in the model structure. In addition, for finding the optimal maintenance interval it is necessary to make trade-offs between several factors, which may conflicting each other as well. The study comprises of mathematical formulating an optimal interval problem based on Multi-Attribute Utility Theory (MAUT). The aim of the proposed research is to develop a methodology with supporting tools for determination of optimal inspection in a maintenance planning to assure and preserve a desired level of performance measure such as reliability, availability, risk, etc. For verification and validation purposes, the proposed methodology (analysis approach) and tools (models) will be applied in a real case which given by the literature.

Keywords Maintenance optimization · Multi-attribute utility theory · Reliability · Availability · Risk

A.H.S. Garmabaki (✉) · A. Ahmadi
Division of Operation and Maintenance Engineering, Luleå University
of Technology, Luleå, Sweden
e-mail: amir.garmabaki@ltu.se; garmabaki@gmail.com

A. Ahmadi
e-mail: alireza.ahmadi@ltu.se

A.H.S. Garmabaki
Department of Mathematics and Computer Science, Islamic Azad University,
Nour Branch, Nour, Iran

M. Ahmadi
Department of Industrial Engineering, Mazandaran University of Science
and Technology, Babol, Iran
e-mail: mahdiehahmadi@yahoo.com

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_2

1 Introduction

Multi-criteria decision making (MCDM) is one of the most well-known branches of decision making. According to many authors (see, for instance, [1]) MCDM is divided into multi-objective decision making (MODM) and multi-attribute decision making (MADM). MCDM is concerned with the methods and procedure by which multiple criteria can be formally incorporated into the analytical process [2]. There are several methods proposed by literature. The weighted sum model (WSM) is the earliest and probably the most widely used methods. The weighted product model (WPM) can be considered as a modification of the WSM, and has been proposed in order to overcome some of its weaknesses. The analytic hierarchy process (AHP), as proposed by Saaty [3], is a later development and it has recently become increasingly popular in different area. Belton and Gear [4] modified AHP method and the new approach is more consistent than the original AHP. Some other popular methods proposed by literature are the VIKOR and the TOPSIS methods. These methods are based on an aggregating function representing "closeness to the ideal, which originated in the compromise programming method". Both TOPSIS and VIKOR are based on the calculation of distances from the Positive Ideal Solution (PIS) and the Negative Ideal Solution (NIS). Chu et al. [5] are in favors of using VIKOR when there are a larger number of decision makers (DM), and otherwise they recommend the use of TOPSIS. Recently, Ahmadi et al. [6] show that application of the combined AHP, TOPSIS, and VIKOR methodologies are applicable and verified the proposed methodology through a case study for an aircraft system.

Maintenance decision making is a complex task and may take place in several contexts with different types of systems in terms of technology, repairability, reliability and availability requirements, etc. For optimal time determination of the maintenance plan, maintenance management may present scenarios, including several objectives which often competing or conflicting with each other. The objectives can be represented by a set of appropriate measures or attributes, which are used to represent system characteristics. Here, the decision maker not only required to choose the best solution among alternatives, but also have to trade-off between the objectives.

Kralj and Petrovic [7] used multiple objective function to tackle costs and reliability in preventive maintenance. In another study, an optimal interval for preventive maintenance was obtained based on the PROMETHEE method [8]. Gopalaswamy et al. [9] argued for strict selection and lexicographical approaches applied to preventive maintenance, taking into account criteria such as costs, availability and reliability. Most research on preventive maintenance problems in the literature is based on a multi-criteria approach to analyze particular problems using multi-criteria approaches that do not incorporate the most useful advantage of multi attribute utility theory (MAUT). However, some decision models for maintenance are based in MAUT. See [10–12].

Here, we propose an optimal maintenance inspection model based on MAUT. In order to determine optimal time, different criteria such as cost, reliability, and availability are considered in the model framework. In order to provide insight into the problem, a utility function is assessed for each of the relevant objectives. This allows for an appropriate multiple objective utility functions that are used to identify tradeoffs and compare the various objectives in a consistent manner. The basis of utility theory and its underlying quantitative axioms were initially established by Keeney and Raiffa [13]. The decision model has been applied on a real case in an electric power company. The decision level and weight parameter are selected, subjectively and sensitivity analysis is conducted to identify the most sensitive parameter.

The rest of the paper is organized as follows. The proposed model based on MAUT is discussed in Sect. 2. Section 3 shows the numerical example and verified the proposed methodology through a real case study. In addition, a sensitivity analysis is discussed in Sect. 4. Finally, conclusions are given in Sect. 5.

2 Multi Attribute Utility Theory

Multi-attribute utility theory (MAUT) [13] is concerned with expressing the utilities of multiple-attribute outcomes or consequences as a function of the utilities of each attribute taken singly. This approach has been used for choosing the most “desirable alternative” (or project) among many different alternatives. It has been used in a broad range of fields including energy, manufacturing and services, public policy, health care, etc. MAUT can help in these situations by creating a decision model through the elicitation process of expert practitioners.

The theory specifies several possible functions (additive, multiplicative and multi-linear) and the conditions (independence conditions to be met) under which each would be appropriate. As a practical matter, Keeney and Raiffa [13] suggest that for four or more attributes the reasonable models are the additive and the multiplicative. Since our problem contains less than four attributes, we restrict our attention to the additive form. The MAUT process provides a framework through which multiple objectives and uncertainty can be combined to aid managers in making decisions. In order to create a MAUF Problem, single utility functions must be assessed for every identified objective. In our case, we have identified three separate attribute. The objective list utilized for this preliminary analysis is minimization of cost and maximization of reliability and availability. Generally, a MAUF is defined as:

$$\begin{aligned} U(x_1, x_2, \ldots, x_n) &= f[u_1(x_1), u_2(x_2), \ldots, u_n(x_n)] \\ &= \sum_{i=1}^{n} w_i.\, u_i(x_i) \end{aligned} \tag{1}$$

where, $\sum_{i=1}^{n} w_i = 1$

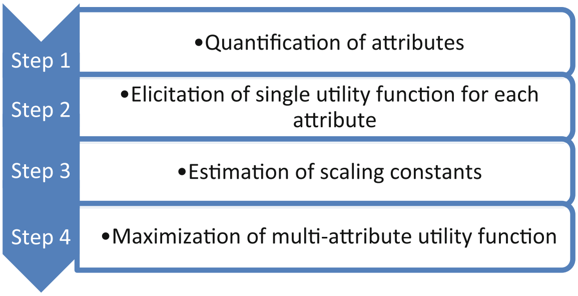

Step 1
•Quantification of attributes
Step 2
•Elicitation of single utility function for each attribute
Step 3
•Estimation of scaling constants
Step 4
•Maximization of multi-attribute utility function

Fig. 1 The structure of MAUT for the determination of optimal inspection time

where, U is a multi-attribute utility function over all utility functions; $u_i(x_i)$ is a single utility function measuring the utility of attribute i; x_i is level of ith attribute. w_i represent the relative importance weights for the utilities. By maximizing the multi-attribute utility function, the best alternative is obtained, under which the attractiveness of the conjoint outcome of attributes is optimized. The main reason for the selection of MAUT in our problem is that scenarios of management can be appropriately represented by the structure of this technique. Furthermore, MAUT has strong theoretical foundations based on the expected utility theory.

In order to obtain structure for utility functions, first we need to make assumptions regarding utility independence and the additive independence. The procedure of the use of it in our problem is discussed in detail by [13]. The utility functions are assessed in the following four steps [13, 14] (Fig. 1).

2.1 Quantification of Attributes

In our case study, cost, availability and reliability are selected as attribute to find out the optimal maintenance policy. The attributes and their mathematical structure are discussed in following subsection.

2.1.1 Cost Modeling

In the preventive replacement age policy subject to breakdown, instead of making a preventive replacement at fix time interval T, the preventive replacement depends on the age of the item. In addition, failure replacement is performed if the system fails before T and the time clock is reset to zero, see [15] for more details. The average cost per unit time based on optimal preventive replacement is given by:

$$\begin{aligned} C(T) &= \frac{c_p R(T) + c_f(1 - R(T))}{T.R(T) + M(T).(1 - R(T))} \\ &= \frac{c_p R(T) + c_f F(T)}{T.R(T) + M(T).F(T)} \end{aligned} \tag{2}$$

where $M(T) = \int_{-\infty}^{T} \frac{tf(t)}{(1-R(T))} dt$
and T is a replacement age at which a preventive replacement takes place, c_p and c_f ($c_f > c_p$) are the cost of a preventive and failure replacement. In both cases, replacement cost includes all costs resulting from the failure and its replacement.

In this model, the numerator equals to the total expected cost per cycle and the denominator equals to the expected cycle length; $F(t)$ and $R(t)$ are the cumulative distribution and reliability functions, respectively. The optimal value of T corresponds to the minimum cost, $C(T)$, can be derived by the first derivation of $C(T)$ with respect to T. This model is discussed in details by Jardine and Tsang [16].

Cost Attribute

The average cost per unit time given by Eq. (2) has a unique minimum C_{Min} which occurs at T_C. Since small value of $C(T)$ is preferred, we define the cost attribute function as:

$$U_{\text{Cost}} = \frac{C_{Min}}{C(T)} \tag{3}$$

2.1.2 Availability Modeling

Availability is defined as the long run probability of the system being available for use at any point in time [17]. This is expressed as a point estimate and calculated from the mean delay and reliability point estimates. There are several different forms of steady state availability depending on the definition of uptime and downtime. The Inherent availability is most common definition in the literature:

$$A_I = \frac{MTTF}{MTTF + MRT} \tag{4}$$

where MRT is the mean repair time and $MTTF$ is the mean time-to-failure.

In our decision problem, optimal preventive replacement age policy subject to breakdown are considered. For above standard definition, the following structure can be derived for single unit.

$$A(T) = \frac{\int_0^T R(t)dt}{\int_0^T R(t)dt + t_p R(T) + t_f(1 - R(T))} \tag{5}$$

where t_p and t_f are the require time of performing a preventive and a failure replacement, respectively. A large value of $A(T)$ is preferred.

Availability Attribute

The average availability per unit time given by Eq. (5) has a unique maximum A_{Max} which occurs at T_A. Since a large value of $A(T)$ is preferred, the availability attribute may be define as:

$$U_{Ava} = \frac{A(T)}{A_{Max}} \tag{6}$$

2.1.3 Reliability Modeling

Reliability is closely associated with the quality of the product. This criteria is one of the main concerns during different stage of product development such as design, testing and operation. Reliability is defined as probability that a system will function over the time period. Reliability can be expressed as

$$\begin{aligned} R(t) &= \Pr(T \geq t) \\ R(t) &= 1 - F(t) \end{aligned} \tag{7}$$

where $R(t) \geq 0, R(0) = 1$ and $\lim_{t \to \infty} R(t) = 0$.

Reliability Attribute

The reliability level of the product at time T, is depend to failure distribution and the interval which is our aim to study. Reliability per unit time given by Eq. (7), has a unique maximum R_{Max} which occurs at T_R. Since a large value of $R(T)$ is preferred, the reliability attribute is given by:

$$U_{\mathrm{Rel}} = \frac{R(T)}{R_{Max}} \tag{8}$$

2.2 Elicitation of Single Utility Function for Each Attribute

The single utility function for each attribute represents management's satisfaction level towards the performance of each attribute. It is usually assessed by a few particular points on the utility curve [18, 19].

More specifically, suppose that the best and worst values of availability are selected first as A^B and A^W. At these boundary points, we have $U(A^W) = 0$ and $U(A^B) = 1$. For cost utility function, highest and lowest budget consumption requirement values are selected as C^W *and* C^B, respectively. Also, at these boundary points, we have $U(C^W) = 0\, and\ U(C^B) = 1$.

To elicit the single utility function the exponential or linear function, may suggested for each attribute given by Eq (9).

$$\begin{cases} U(x) = k_1 x + k_2 & \textit{Linear function} \\ U(x) = k3.\exp\left(-\frac{k_4}{x}\right) & \textit{Exponential function} \end{cases} \tag{9}$$

where k_i are constants which secure $U(\mathrm{x_i}) \in [0, 1]$. Unknown parameter for utility functions,$U(A)$, $U(R)$ and $U(C)$ can be obtained using linear (exponential) form of single utility function with the help of boundary conditions.

The linear utility function is applied for availability and cost attribute. The linear function is applicable when the DM is risk neutral [13]. That is, the DM is neither risk prone nor risk averse. For reliability, the logistic utility function is found to be suitable. This function presents a risk aversion for higher values of R and prone risk for lower values of R, which is the DM's risk behavior for increasing utility function.

2.3 Estimation of Scaling Constants

The following step is the estimation of the scaling constants w_A,w_R and w_C. They indicate the importance weights that management team allocates for each attribute [18, 20]. There are two common methods to assess the scaling constants:

1. Certainty scaling and
2. Probabilistic scaling

Given that the number of attributes considered in our problem is three and we will use probabilistic scaling technique.

Consider three attributes A, R and C as availability, reliability and cost. Let $(A^B, R^B,\ C^B)$ and $(A^W,\ R^W,\ B^W)$ denote the best and worst possible consequence, respectively (Fig. 2). There is a certain joint outcome $(A^B, R^B,\ C^W)$ comprised three attribute *A, R* and *C* at the best and worst level with probability *p* and (*1*−*p*), respectively. In these situations, the weight for attribute *C* equals *p*, where *p* is the indifference probability between them, see [18].

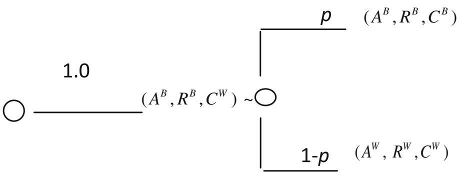


Fig. 2 Assessing scaling constants

2.4 Maximization of Multi-attribute Utility Function

Based on the previously estimated single utility functions and scaling constants, the additive form of the multi-attribute utility function in our problem can be obtained. That is

$$\begin{aligned} Max:\ \ U(A, R, C) = w_A \times \ U(A) + w_R \times U(R) + w_C \times \ U(C) \\ w_A + w_R + w_C = 1 \end{aligned} \tag{10}$$

where $\mathrm{w_A}$, w_R and w_C are the weight parameters for attribute A, R and C, respectively. $U(A), U(R)$ and $U(C)$ are the single utility function for availability, reliability and cost attribute. It may note that the $U(A, R, C)$ function is *Maximum* type and it has been written in terms of A, R and C.. By maximizing this multi-attribute utility function, the optimal inspection, T^* will be obtained. It is worth noting here that the additive form of multi-attribute utility function is based on the utility independence and the additive independence assumptions.

3 Numerical Example

This numerical application is conducted to verify MAUT in maintenance application. Assume that 2-parameter Weibull model is selected as failure distribution which are given by Eq (11) and the parameter of the model and attributes are given in Table 1.

Table 1 Estimated parameter from real application [21]

β	3	Shape parameter
η	1200	Scale parameter
t_p	0.2	Time of performing preventive maintenance
t_f	0.4	Time of performing corrective maintenance
c_p	600	Cost of preventive maintenance
c_f	1200	Cost of corrective maintenance

$$F(T) = 1 - \exp(-(T/\eta)^{\beta})$$
$$f(T) = \frac{\beta}{\eta}.\left(\frac{T}{\eta}\right)^{\beta-1}.\exp(-(T/\eta)^{\beta}) \tag{11}$$

In addition, the best and worse level for each attribute are given in Table 2. The linear utility function is applied for availability and cost attribute. In addition, the logistic utility function is considered for reliability attribute. For each attribute, the constant coefficients are calculated and given in Table 2. The availability, reliability and cost attribute are plotted in Figs. 3, 4 and 5.

Table 2 Attributes function and coefficients

Attributes	Best	Worse	Function	Coefficient value
Availability attribute	$A^B = 0.95$	$A^W = 0.25$	$U(x) = k_1A + k_2$	$k_1 = 1.428;$ $k_2 = -0.357$
Reliability attribute	$R^B = 0.9$	$R^W = 0.3$	$U(x) = k3.\exp\left(-\frac{k_4}{R}\right)$	$k_3 = 9.985;$ $k_4 = 2.0718$
Cost attribute	$C^B = 0.35$	$C^W = 1$	$U(x) = k_5A + k_6$	$k_5 = 1.5384;$ $k_6 = -0.538$

Fig. 3 The availability attribute

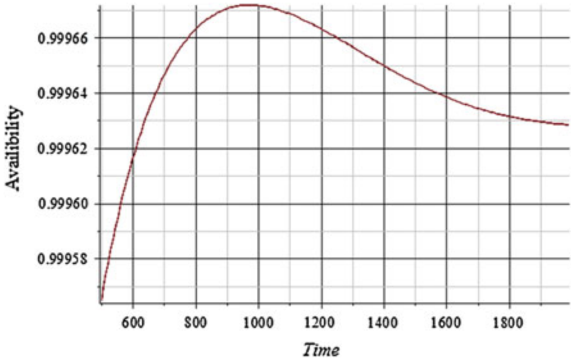


Fig. 4 The reliability attribute

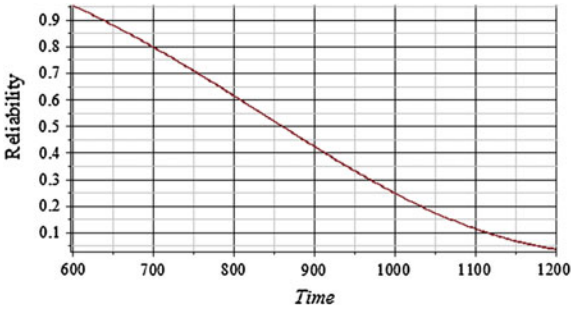


Fig. 5 The cost attribute

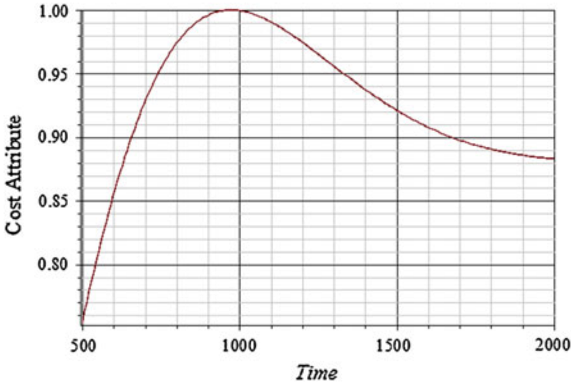
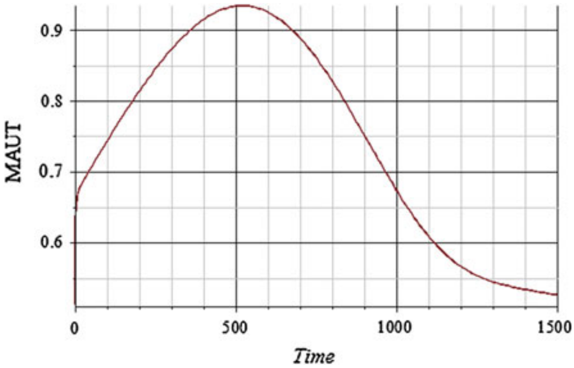


Fig. 6 Multi attribute utility function



The behavior of MAUF function is given in Fig. 6. The optimal inspection time by considering three attribute with above weight occur at $t \in [490, 550]$. More specifically, when we consider only cost for determination of optimal inspection time, we get $t = 950$ which seems is more delay for inspection time.

4 Sensitivity Analysis of the Model Parameters

From the discussion given in the preceding section, it is good to know that the optimal decision-making depends on various parameters that may not be precise.

The use of sensitivity analysis will help the analyst to understand how changing the parameters of the model will affect the decision outcome. The decision model is then rerun by holding all other parameters constant. We have conducted sensitivity analysis by calculating the relative change of optimal time based different parameters given in Table 3. The sensitivity of the optimal inspection time with respect to

Table 3 Sensitivity analysis results based on model parameter

	$p\%$					
$\varDelta^{u^*}_{\rho,\theta}$	−30 %	−20 %	−10 %	10 %	20 %	30 %
$\varDelta^{u^*}_{\rho,A^B}$	30 %	10 %	1 %	NA	NA	NA
$\varDelta^{u^*}_{\rho,A^W}$	4 %	2 %	1 %	1 %	1 %	2 %
$\varDelta^{u^*}_{\rho,C^B}$	6.5 %	6 %	5 %	NA	NA	NA
$\varDelta^{u^*}_{\rho,C^W}$	8 %	7 %	4 %	2 %	2 %	2 %
$\varDelta^{u^*}_{\rho,R^B}$	NF	NF	NF	10 %	NA	NA

Note: Na, impossible change; Nf, infeasible solution

a model parameter, can be quantified by $\varDelta^{u^*}_{\rho,\theta}$, which are the relative changes of optimal utility level, $u^*(\theta)$ when θ is changed by 100p%, i.e.,

$$\varDelta^{u^*}_{\rho,\theta} = \left| \frac{u^*(\theta + p\theta) - u^*(\theta)}{u^*(\theta)} \right| \tag{12}$$

In addition, different weight are assign to the attribute and the results are plotted in Fig. 7. The values of different weight are given in Table 4.

It can be seen that the sensitivity of optimal interval with respect to model parameters A^W and positive effect of C^W is at acceptably low levels, e.g., when $A^W(C^W)$ increases by 30 % (decreases by −30 %) the relative changes in $\varDelta$ are 2 and 4 %, respectively. Results in Table 3 reveal that A^B and negative part of C^B and C^W are slightly more sensitive parameter than other parameters.

In addition, negative change of w_R did not reveal the high level of sensitivity and positive effect of w_R will reduce inspection time.

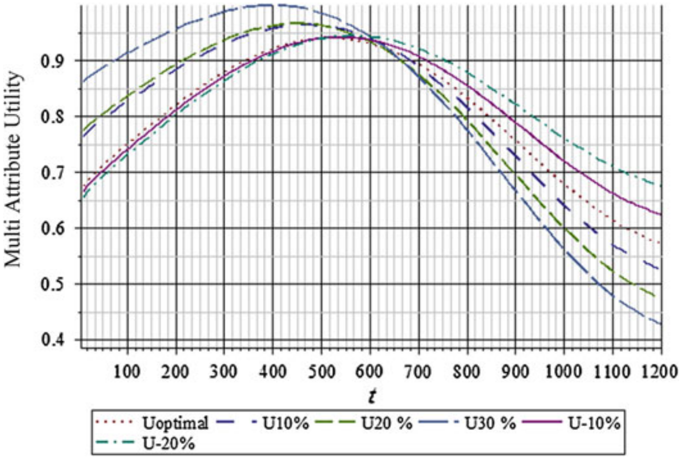


Fig. 7 Sensitivity analysis on weight parameters

Table 4 Different weight of w_R in optimization problem

	$U_{-20\,\%}$	$U_{-10\,\%}$	U_{Optimal}	$U_{10\,\%}$	$U_{20\,\%}$	$U_{30\,\%}$
w_R	0.35	0.4	0.45	0.5	0.55	0.6
w_A	0.35	0.3	0.25	0.25	0.2	0.2
w_C	0.3	0.3	0.3	0.25	0.25	0.2

5 Conclusion

In this paper, we have developed a multi attribute utility model for the preventive replacement age policy subject to breakdown. Reliability, availability and cost are considered as three main attribute in our decision problem. By using MAUT, it is possible to make trade-offs between several factors, which may conflicting each other as well. In addition, the optimal solution depends not only on the failure distribution and the cost ratio, but also on the maintenance time ratio as well as the relative importance of the attributes. The MAUT is important for the maintenance and reliability community when a context of service production systems is to be taken into account due to disturbances caused by failures in the system. A numerical application has illustrated the use of the decision model and the procedure.

References

1. Zimmermann H (1992) Fuzzy set theory and its applications , 2nd, Revised Edition, Springer, Berlin
2. Steuer R (1986) Multiple criteria optimization: theory, computation, and application
3. Saaty TL (1980) The analytic hierarchy process, McGraw-Hill, New York
4. Belton V, Gear T (1983) On a short-coming of Saaty's method of analytic hierarchies. Omega 11:228–230
5. Chu M-T, Shyu J, Tzeng G-H, Khosla R (2007) Comparison among three analytical methods for knowledge communities group-decision analysis. Expert Syst Appl 33:1011–1024
6. Ahmadi A, Gupta S, Karim R, Kumar U (2010) Selection of maintenance strategy for aircraft systems using multi-criteria decision making methodologies. Int J Reliab Qual Saf Eng 17:223–243
7. Kralj B, Petrovic R (1995) A multiobjective optimization approach to thermal generating units maintenance scheduling. Eur J Oper Res 84:481–493
8. Chareonsuk C, Nagarur N, Tabucanon MT (1997) A multicriteria approach to the selection of preventive maintenance intervals. Int J Prod Econ 49:55–64
9. Gopalaswamy V, Rice J, Miller FG (1993) Transit vehicle component maintenance policy via multiple criteria decision making methods. J Oper Res Soc 44:37–50
10. de Almeida AT, de Souza FMC (1993) Decision theory in maintenance strategy for a 2-unit redundant standby system. IEEE Trans Reliab 42:401–407
11. Jiang R, Ji P (2002) Age replacement policy: a multi-attribute value model. Reliab Eng Syst Saf 76:311–318
12. Ferreira RJ, de Almeida AT, Cavalcante CA (2009) A multi-criteria decision model to determine inspection intervals of condition monitoring based on delay time analysis. Reliab Eng Syst Saf 94:905–912

13. Keeney RL, Raiffa H (1993) Decisions with multiple objectives: preferences and value trade-offs. Cambridge university press, Cambridge
14. Kapur PK, Garmabaki AHS, Singh JNP (2012) “The Optimal Time of New Generation Product in the Market. Commun Dependability Qual Manage Int J 15:123–137
15. Smith WL (1955) Regenerative stochastic processes. Proc R Soc London Ser A Math Phys Sci 232:6–31, 1955
16. Jardine AKS, Tsang AHC (2006) Maintenance, replacement, and reliability: theory and applications. CRC/Taylor & Francis, Boca Raton
17. Billinton R, Allan RN (1992) Reliability evaluation of engineering systems. Springer, Berlin
18. Keeney RL, Raiffa H (1976) Decisions with multiple objectives: preferences and value tradeoffs. Wiley, New York
19. Garmabaki AHS, Aggarwal AG, Kapur PK, Yadavali VSS (2012) Modeling two-dimensional software multi-upgradation and related release problem (A Multi-Attribute Utility Approach). Int J Reliab Qual Saf Eng 19:1250012
20. Neumann JV, Morgenstern O (1947) Theory of games and economic behavior, 2nd edn. Princeton University Press, Princeton
21. Almeida AT (2012) Multicriteria model for selection of preventive maintenance intervals. Qual Reliab Eng Int 28:585–593

Optimum Proactive Maintenance for Critical Infrastructures Subject to Multiple Degradation and Environmental Shocks

Mahmood Shafiee

Abstract Critical infrastructures (e.g. power networks, transport systems, financial services and telecommunication) constitute the backbone of the society. A failure in these systems may result in substantial costs in terms of lost service delivery and emergency maintenance operations. Failures of critical infrastructures mainly occur as a result of various degradation (deterioration) processes in their consisting units as well as due to external shocks arising from surrounding environment. In order to avoid such failures, various *proactive* maintenance policies, including routine inspection, age (usage)-based replacement, and condition-based maintenance are commonly applied. In this paper, we formulate an optimization framework for proactive maintenance planning of critical infrastructures subjected to stochastic degradation and environmental damages. The infrastructure in our study is composed of multiple identical sub-systems, each exposed to a gradual degradation phenomenon. The environmental shocks are divided into two types of minor (with probability p) and major (with probability $1-p$, where $0 \leq p \leq 1$). A minor shock causes a disruption in system operation without resulting in any failure, while a major shock stops the system and requires a costly replacement. The performance of the proposed maintenance policies, regarding the objective of minimum average long-run maintenance cost per unit time are compared to existing practices of maintenance. Several case studies within the subsea pipeline, marine renewable energy, and the rail transport industries are presented to illustrate the results.

Keywords Critical infrastructure · Proactive maintenance, stochastic degradation, environmental shock · Repair, replacement

M. Shafiee (✉)
Cranfield University, College Road, Cranfield, Bedfordshire MK43 0AL, UK
e-mail: m.shafiee@cranfield.ac.uk

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_3

1 Introduction

Critical infrastructures, such as power networks, transport systems, financial services and telecommunication constitute the backbone of the society. These systems provide services that are important in maintaining the essential functions of society. A failure in these systems may result in substantial costs in terms of lost service delivery and emergency maintenance operations. For this reason, the requirement for improving the reliability of critical infrastructures has recently experienced a great increase. In December 2006, the European Commission approved a programme for critical infrastructure protection (CIP) aiming to identify and protect the EU's critical infrastructures in case of faults, incidents or attacks (for more see [1]).

Failures of critical infrastructures mainly occur as a result of various degradation (deterioration) processes in their constituent sub-systems as well as due to external shocks arising from surrounding environment. Degradation is a complex multi-dimensional process which depends on numerous physical and mechanical factors (e.g. material, stress loads) and is manifested in different forms of wear, fatigue, and crack generation [2]. Any of these forms or their combination can result in a failure if their length reaches a critical level. In this case, the system undergoes an unplanned maintenance action which includes performing a replacement on the failed item. On the other side, environmental shocks are divided into two types of minor (with probability p) and major (with probability $1-p$, where $0 \leq p \leq 1$) [3]. A minor shock causes a disruption in system operation without resulting in any failure, while a major shock stops the system and requires a costly replacement. So, it is crucial to continuously monitor and evaluate the degradation state and operating condition of critical assets so that unexpected failures can be eliminated (minimized).

Currently, a large number of sensors and control devices are installed at various locations of system networks to collect condition data (e.g. temperature data, deterioration modes and causes, fatigue cracks size, damage propagation). The collected information is frequently transferred to supervisory control and data acquisition (SCADA) system and is stored in databases. The system analysts use the SCADA database to schedule the inspection and maintenance tasks when required. In this regard, various *proactive* maintenance policies, including routine inspection, age/usage-based replacement, and condition-based maintenance are commonly applied to critical infrastructures protection [4]. A brief review of the literature shows that a lot of research has been done on optimization of proactive maintenance policies for isolated infrastructures (or being possibly assimilated to single-infrastructure systems). However, there often exist strong *correlations* among the failure modes as well as between the sub-systems of various infrastructures [5]. Neglecting these correlations while optimizing maintenance policies leads to sub-optimal or even wrong solutions to the problem and thereby, increased cost of maintenance and system downtime.

In this paper, we formulate an optimization framework for proactive maintenance planning of critical infrastructures subjected to stochastic degradation and environmental damages. The infrastructure in our study is composed of multiple identical sub-systems (e.g. a subsea pipeline with multiple identical pipe segments), each exposed to a gradual degradation phenomenon. The system undergoes a proactive maintenance action according to either one of the following schemes:

(i) POLICY I: An *age-dependent maintenance action* is carried out at fixed time intervals *k*T (*k* = 1, 2, …) after the installation (see Fig. 1a).

(ii) POLICY II: A *degradation-dependent maintenance action* is carried out when the condition signal in a sub-system reaches an alert threshold *d* (smaller than fault threshold D) (see Fig. 1b). In order to take the advantage of system dependence, a preventive repair action is also performed on other safe sub-systems.

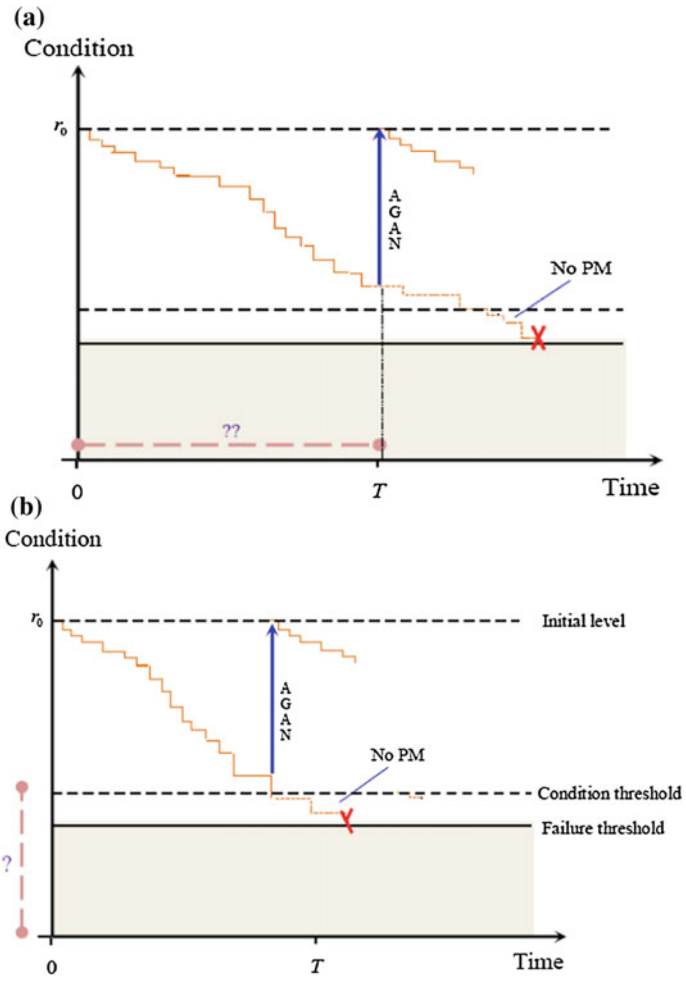


Fig. 1 The proposed proactive maintenance policy

(iii) POLICY III: A *proactive maintenance action* is conducted at fixed time intervals kT or when the condition signal in a sub-system exceeds an alert threshold d, whichever comes first.

The problem is to find out the optimum block replacement time T* (>0) and/or condition threshold d* (<D) such that the system's average long-run maintenance cost per unit time is minimized. Our objective function includes all costs due to corrective replacement, preventive maintenance and repairs, and loss of service delivery. The explicit expression of the objective function is derived and under certain conditions, the existence and uniqueness of the optimal solution are shown. The performance of the proposed maintenance policies are evaluated using a Monte-Carlo simulation technique and are compared to existing practices of maintenance.

The rest of this paper is organized as follows. In Sect. 2, we present the problem definition. In Sect. 3, we construct our optimization framework and discuss the properties of the optimal solution. Several case studies in the subsea pipeline, marine renewable energy and the rail transport industries are presented in Sect. 4.

2 Problem Definition

Notation

n	number of sub-systems in the infrastructure
i	index for sub-systems; $i \in \{1, 2, \ldots, n\}$
$m(t)$ $[M(t)]$	intensity [mean value] function of degradation process in a sub-system
j	index for number of degradation processes
T_{ij}	initiation time of the jth degradation process in the sub-system i
$\overline{F}_{T_{ij}}(.)$	survival function of T_{ij}
$X_{ij}(t)$	level of the jth degradation process in the sub-system i at time point t after initiation
U_{ij}^{x}	length of the interval between the initiation time of the jth degradation process in the sub-system i to the time that it attains a size x
$g_{U^*}(\cdot)[G_{U_{ij}^{x}}(\cdot)]$	probability density [cumulative distribution] function of U_{ij}^{x}
$\alpha[\beta]$	shape [scale] parameter of the gamma distribution
$\Gamma(\bullet)[\gamma(.,.)]$	gamma [incomplete gamma] function
S_{ij}^{x}	time (since $t = 0$) to attain size x for the jth degradation process in the sub-system i
$h(.)[H(.)]$	intensity [mean value] function of environmental shocks
p $[1\!-\!p]$	probability that an environmental shock is catastrophic [minor]; $0 \le p \le 1$

T_f	time to arrive a catastrophic shock
$\overline{F}_{T_f}(.)$	survival function of T_f
D	fault threshold of degradation for sub-systems
d (<D)	control threshold for PM
S_d	time point that, for the first time, the degradation level of a sub-system reaches d
$\overline{F}_{S_d}(.)[\theta_d(.)]$	survival [hazard rate] function of S_d
T	PM interval
C_T	fixed cost of a planned PM
$C(d)$	cost of performing a major repair for a sub-system with degradation level d
C_R	fixed cost of replacing a failed sub-system
c_m	expected cost of service loss due to a minor shock
C_0	set-up cost for a planned PM action at time T
C_1	set-up cost for an unplanned major repair action at threshold d
C_2	set-up cost for a corrective replacement action
$C_n(d, T)$	average long-run maintenance cost for a sub-system per unit time

Initiation of degradation—Consider a critical infrastructure which is composed of n identical sub-systems connected in series (in terms of reliability) and all working independently of each other. A failure of the sub-system j (= 1, 2, …, n) causes the failure of entire system, which is immediately detected. Each sub-system is subject to a random number of degradation processes independently from the others. Suppose that the degradation processes in the sub-system i are initiated by point events that follow a non-homogeneous Poisson process (NHPP), $\{N_{1i}(t) \equiv N_1(t); t \geq 0\}$ with intensity function (rate) $m(t)$ and mean value function $M(t)$, i.e., [6]

$$M(t) = \int_0^t m(y)dy,\ t \geq 0. \tag{1}$$

Let T_{ij}, $i = 1, 2, \ldots, n, j = 1, 2, \ldots$, denote the initiation time of the jth degradation process in the sub-system i. Then, the survival function that corresponds to the random variable T_{ij} is given by

$$\overline{F}_{T_{ij}}(t) \equiv \overline{F}_{T_j}(t) = P\{N_1(t) < j\} = e^{-M(t)} \times \sum\nolimits_{k=0}^{j-1} \frac{[M(t)]^k}{k!}. \tag{2}$$

Propagation of degradation—Assume that all degradation processes in the sub-systems propagate independently from each other. Let $X_{ij}(t)$, $i = 1, 2, \ldots, n$, $j = 1,2,\ldots$, be the level of the jth degradation process in the sub-system i at time point t after initiation. Thus, $X_{ij}(t)$, are the increasing stochastic processes of degradation. Denote by U_{ij}^x the length of the time interval between the initiation time of the jth degradation process in the sub-system i to the time that it attains a size x (the first passage time). Let $X_{ij}(t) \equiv X(t)$ and $U_{ij}^{x} \equiv U^{x}$, which means that the initiated degradation processes are statistically identical for different initiating events affecting the sub-systems. We also assume that the corresponding stochastic processes are independent. Thus,

$$U^{x} = \inf\{t \geq 0 : X(t) \geq x\}, x > 0. \tag{3}$$

In this paper, we model the level of a degradation process using the stochastic *gamma* process. Assume that $X(t)$ is a homogeneous gamma process with shape and scale parameters αt and β, respectively. Thus, the density and the cumulative distribution function of U^{x} are given respectively by [7]

$$g_{U^x}(t) = \frac{\beta^{\alpha t}}{\Gamma(\alpha t)} x^{\alpha t - 1} e^{-\beta x},\ t \geq 0,\ \alpha,\ \beta > 0, \tag{4}$$

$$G_{U^x}(t) = \frac{\gamma(\alpha t,\ \beta x)}{\Gamma(\alpha t)},\ t \geq 0,\ \alpha,\ \beta > 0, \tag{5}$$

where $\Gamma(\cdot)[\gamma(.,.)]$ denotes the gamma [incomplete gamma] function, i.e.,

$$\Gamma(\upsilon) = \int_0^\infty z^{\upsilon-1} e^{-z} dz;\ \gamma(\upsilon,\ u) = \int_u^\infty z^{\upsilon-1} e^{-z} dz,\ \upsilon,\ u > 0. \tag{6}$$

Denote by S_{ij}^{x} the time point (since $t = 0$) when the level of the jth degradation process in the sub-system i exceeds x. Then,

$$S_{ij}^{x} = T_{ij} + U^{x},\ x > 0,\ i = 1,\ 2,\ \ldots, n,\ j = 1,\ 2,\ \ldots. \tag{7}$$

Let $\{N_{Si}^{x}(t);\ t \geq 0\}$, $i = 1,\ 2,\ \ldots,\ n,$ be the counting process associated with the random variable S_{ij}^{x}, where $N_{Si}^{x}(t)$ denotes the total number of degradation processes in the sub-system i that exceeds a size x in the interval $[0,\ t)$. Then, we can show that $\{N_{Si}^{x}(t);\ t \geq 0\}$ is an NHPP with intensity function,

$$\theta_x(t) = \int_0^t m\,(t) g_{U^*}(t - y) dy \equiv m\,(t) * g_{U^*}(t),\ x > 0, \tag{8}$$

where the symbol * represents convolution function and $g_{U}^{x}(.)$ is given by Eq. (4).

Environmental shocks—Suppose that the environmental shocks arrive at the whole infrastructure according to a non-homogeneous Poisson process (NHPP) $\{N_2(t);\ t \geq 0\}$ with intensity function $h(t)$ and mean value function $H(t)$, i.e.,

$$H(t) = \int_0^t h\,(t) dy,\ t \geq 0. \tag{9}$$

External shocks are minor with probability $1-p$ and catastrophic with probability p $(0 \leq p \leq 1)$. We denote by T_f the time of arrival of a catastrophic shock. Then, the survival function of the random variable T_f is given by

$$\overline{F}_{T_f}(t) = \exp\{-pH(t)\},\ 0 \leq p \leq 1. \tag{10}$$

3 Maintenance Optimization

Let S_d denote the time that, for the first time, a degradation process in one of the sub-systems exceeds the threshold d, i.e.,

$$S_d = \min\left\{S_{ij}^d,\ i = 1,\ 2,\ \ldots, n,\ j = 1,\ 2,\ \ldots\right\},\ 0 < d \leq D, \tag{11}$$

where S_{ij}^d is the time to attain size d for the jth degradation process in the sub-system i. Then, taking into account Eq. (8) for $x = d$, the survival function of S_d can be written as

$$\overline{F}_{S_d}(t) = P\{S_d > t\} = \prod_{i=1}^{n} P\{N_{S_i^d}(t) = 0\} = \exp\{-n\,(m(t) * G_{U^d}(t))\}, \tag{12}$$

where n is the number of sub-systems in an infrastructure, and G_U^d (.) is given by replacing x with d in Eq. (5). Now, let X_r denote the duration of the renewal cycle defined by the time interval between successive maintenance actions. Under the assumptions of the model,

$$X_r = min\left(T,\ S_d,\ T_f\right),\ T > 0,\ 0 < d \leq D, \tag{13}$$

Hence, the expected duration of a renewal cycle, $E(X_r)$ is given by

$$E(X_r) = \int_0^T \overline{F}_{S_d}(t)\,\overline{F}_{T_f}(t)dt, \tag{14}$$

where $\overline{F}_{T_f}(t)$ is given by Eq. (10), and $\overline{F}_{S_d}(\cdot)$ is given by Eq. (12).

Denote by $E[N_{E,m}\ (d, T)]$ the expected number of minor shocks that arrive at the infrastructure during the renewal cycle. Then,

$$E\left[N_{E,m}(d, T)\right] = (1-p)\int_0^T h(t)\,\overline{F}_{S_d}(t)\,\overline{F}_{T_f}(t)dt. \tag{15}$$

The cost of performing a planned PM action and the replacement cost for each sub-system are C_T and C_R, respectively. The cost of performing a major repair at condition threshold d is represented by function $C(d)$, which is a non-negative, non-decreasing differentiable function of d. In addition to the repair or replacement costs, conducting a maintenance task incurs a fixed set-up cost, which usually includes the costs for ordering the spare parts, equipping the maintenance teams, and hiring the maintenance personnel and transport vehicles. We assume that the maintenance set-up costs for a planned PM at time T, a major repair at control threshold d and a replacement task are respectively C_0, C_1, and C_2, where $C_2 \geq C_1 \geq C_0 > 0$. Also, the expected cost of service disruption due to a minor shock is c_m.

Let $S(t)$ represent the expected cost of operating the system for the time interval $[0, t)$. From the *renewal reward theorem* (see [8, p. 52]), the average long-run maintenance cost per unit time is the operational cost incurred in a renewal cycle divided by the length of the expected cycle. Then, the average long-run maintenance cost for a sub-system per unit time, denoted by $C_n(d, T)$ is given by:

$$C_n(d,\, T) = \frac{1}{n} \lim_{t \to \infty} \frac{S(t)}{t}. \tag{16}$$

The average long-run maintenance cost of a sub-system per unit time under the proposed maintenance policies are as follows:

POLICY I:

$$C_n(D,T) = \frac{(C_0 + nC_T) \; + \; \int_0^T \xi_n(D,t)\overline{F}_{S_D}(t)\,\overline{F}_{T_f}(t)\,dt}{n \; \int_0^T \overline{F}_{S_D}(t)\,\overline{F}_{T_f}(t)\,dt},\; T > 0, \tag{17}$$

where $\xi_n(D,\, t)$ is defined as below:

$$\begin{aligned} \xi_n(D,t) = n\,[(C_2 - C_0) + (C_R - C_T)]\,\theta_D(t) \\ + [p\,(C_2 - C_0 + n\,(C_R - C_T)) + (1-p)c_m]\,h(t). \end{aligned} \tag{18}$$

POLICY II:

$$C_n(d,\infty) = \frac{(C_0 + nC_T) \; + \; \int_0^\infty \zeta_n(d,t)\overline{F}_{S_d}(t)\,\overline{F}_{T_f}(t)\,dt}{n \; \int_0^\infty \overline{F}_{S_d}(t)\,\overline{F}_{T_f}(t)\,dt}, 0 < d \leq D. \tag{19}$$

where $\xi_n(d,\, t)$ is defined as below:

$$\begin{aligned} \xi_n(d,t) = n\,[(C_1 - C_0) + (C(d) - C_T)]\,\theta_d(t), \\ + [p\,(C_2 - C_0 + n\,(C_R - C_T)) + (1-p)c_m]\,h(t). \end{aligned} \tag{20}$$

POLICY III:

$$C_n(d,T) = \frac{(C_0 + nC_T) + \int_0^T \xi_n(d,t)\overline{F}_{S_d}(t)\,\overline{F}_{T_f}(t)\,dt}{n\int_0^T \overline{F}_{S_d}(t)\,\overline{F}_{T_f}(t)\,dt}, 0 < d \leq D. \tag{21}$$

Proposition 1 Let T_n° is an optimal solution that minimizes the objective function C_n (D, T) in Eq. (17). Then,

i. If $\xi_n(D, \infty) > C_n(D, \infty)$, there exists a finite T_n° minimizing C_n (D, T).
ii. If $\xi_n(D, T)$ is strictly increasing and $\xi_n(D, \infty) > C_n(D, \infty)$, there exists a unique, finite minimum.
iii. If $\xi_n(D, T)$ is non-decreasing and $\xi_n(D, \infty) \leq C_n(D, \infty) < \infty$, then $T_n^{\circ} \to \infty$ (reactive response maintenance policy).

Proposition 2 Let $m(t)$ and $h(t)$ be two differentiable non-decreasing functions of t, and assume $\lim_{d–0} C(d) \geq C_T$. There exists an optimal solution d_n° that minimizes the function C_n (d, ∞) in Eq. (19) if the function $C(d)\theta_d(t)$ is strictly increasing in d for each t and the derivative of the function $[(C_1 - C_0) + (C(d) - C_T)]\theta_d(t)$ is sufficiently large.

4 Case Studies

In order to illustrate the proposed policies, the model is applied to maintenance of the three below infrastructures:

4.1 A Subsea Pipeline

A 20-inch oil export pipeline which is used to transport oil from offshore platform to an onshore treatment plant was studied [9]. This pipeline operates under a pressure of 800 psig and temperature of 40 °C and is subject to corrosion and current shocks (see Fig. 2a).

4.2 A Wind Turbine Rotor-Blades

A three-bladed offshore wind turbine system subjected to fatigue cracks and wind loads was studied [10]. The wind turbine has a condition monitoring system that measures a wide range of temperature, noise and vibration parameters (see Fig. 2b).

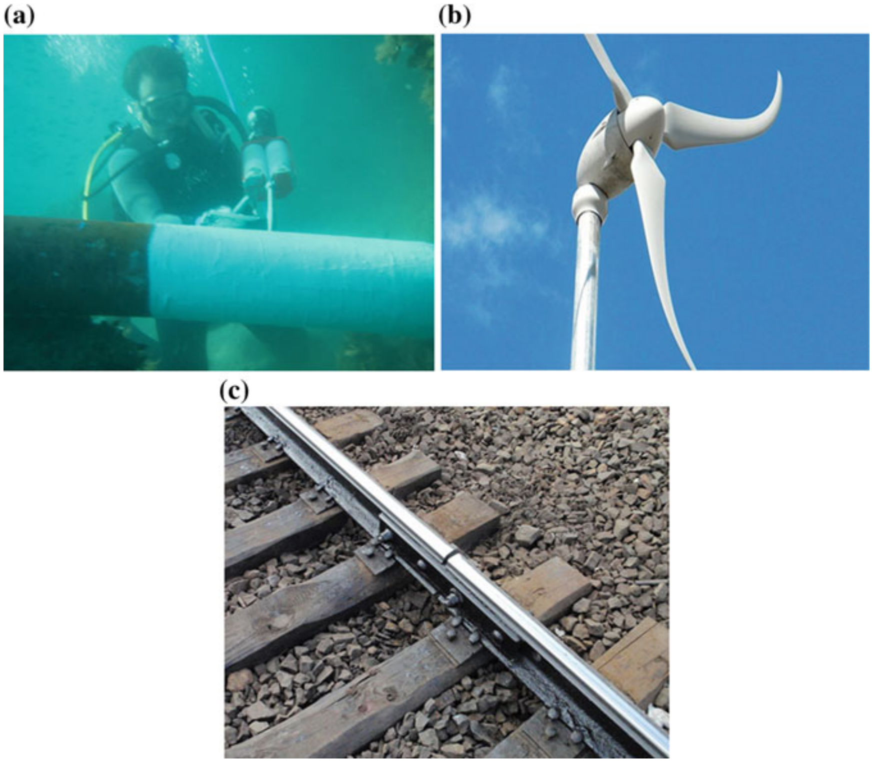


Fig. 2 **a** A subsea pipeline, **b** a three-bladed rotor, **c** a rail track

4.3 A Rail Track

A 60E1 rail track on a small track section subjected to degradation and icing shocks was studied [11] (see Fig. 2c).

References

1. European commission http://ec.europa.eu/index_en.htm
2. Shafiee M, Finkelstein M (2015) An optimal age-based group maintenance policy for multi-unit degrading systems. Reliab Eng Sys Saf 134:230–238
3. Castro IT (2011) A maintenance strategy for systems subject to competing failure modes due to multiple internal defects and external shocks. Proceedings of the European Safety and Reliability Conference. CRC Press, Troyes, pp 770–775
4. Shafiee M, Chukova S (2013) Maintenance models in warranty: A literature review. Eur J Oper Res 229(3):561–572
5. Fiondella L, Liudong X (2015) Discrete and continuous reliability models for systems with identically distributed correlated components. Reliab Eng Sys Saf 133:1–10

6. Finkelstein M (2008) Failure rate modelling for reliability and risk. Springer, London
7. Park C, Padgett WJ (2008) Cumulative damage models based on gamma processes. Encyclopedia of statistics in quality and reliability. doi: 10.1002/9780470061572.eqr119
8. Ross SM (1970) Applied probability models with optimization applications. Holden-Day, San Francisco
9. Shafiee M, Ayudiani PS (2015) Development of a risk-based integrity model for offshore energy infrastructures—application to oil and gas pipelines. Int J Process Sys Eng (in print)
10. Shafiee M, Finkelstein M, Bérenguer C (2015) An opportunistic condition-based maintenance policy for offshore wind turbine blades subjected to degradation and environmental shocks. Reliab Eng Sys Saf 142:463–471
11. Shafiee M, Patriksson M, Chukova S (2014) An optimal age-usage maintenance strategy containing a failure penalty for application to railway tracks. J Rail Rapid Transit. doi: 10.1177/0954409714543337

Risk Informed In-Service Inspection of PWR Nuclear Power Plant Piping Components Subjected to Erosion-Corrosion Using Markov Chain Model

K. Balaji Rao, M.B. Anoop, Gopika Vinod and H.S. Kushwaha

Abstract A Markov Chain (MC) model for failure probability assessment of power plant piping components against erosion-corrosion is proposed. In the MC model, the state space is the degradation state of the system represented by the ratio of the loss in wall thickness due to erosion-corrosion to the original wall thickness of the pipe, and the index space is the time. The use of the proposed MC model is illustrated through an example problem. The model proposed by Abdulsalam and Stanley is used for determining the rate of erosion-corrosion in the example, and, the pipe diameter, pipewall thickness, temperature, pH value, flow velocity, and model error are considered as random variables. From the results obtained, it is noted that there is a need to consider the correlation between degradation at two successive times for obtaining conservative estimates of failure probability against rupture.

1 Introduction

Erosion-Corrosion (EC) is one of the major causes of material degradation of carbon steel piping systems carrying water (single phase) or wet steam (two phase) in Pressurized Heavy Water (PWR) Nuclear Power Plants. The piping systems susceptible to erosion-corrosion damage include feedwater, condensate, extraction steam, turbine exhaust, and, feedwater heater and moisture separator, reheater vents and drains [1]. Significant degradation of pipe wall thickness has been reported in a

K. Balaji Rao (✉) · M.B. Anoop
CSIR-SERC, CSIR Campus, Taramani, Chennai 600 113, India
e-mail: balaji@serc.res.in

G. Vinod
BARC, Homi Bhabha National Institute, Mumbai, India

H.S. Kushwaha
Department of Atomic Energy, Mumbai, India

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_4

number of operating nuclear power plants resulting in fatal accidents, and costly repairs. Hence, an assessment of the resistance degradation based on a suitable wear rate model is essential to predict the life of the piping components against erosion-corrosion damage. The selection of the model for estimation of erosion-corrosion rate should, amongst other factors, be based on its range of applicability and ease of application. Use of such models would help in evolving better strategies of inspection which can be carried out using high precision inspection methods such as radiography, thermography and ultrasonic testing to check the safety of the piping components and replace the susceptible piping components or to carry out the necessary maintenance in time.

For a given piping component (viz. Elbow, Tee) and operating conditions, the EC rate is known to vary [1]. The phenomenon of EC being complex, modeling error also need to be considered. The EC wear rate predicted and modeling error associated with the prediction should be considered as random. The modeling error also accounts for the inherent variations in the phenomenon of EC. In this paper, a Markov Chain (MC) model for failure probability assessment of power plant piping components against erosion-corrosion is proposed. Using the proposed model, the variations in failure probabilities against rupture with time for a power plant piping component are determined. From the results obtained, it is noted that there is a need to consider the correlation between degradation at two successive times for obtaining conservative estimates of failure probability.

2 Modeling Erosion-Corrosion Rate

Erosion-Corrosion is an accelerated form of corrosion caused by the relative motion between corrosive medium (with or without suspended particles) and metal surface leading to loss of material [2]. Modeling erosion-corrosion phenomenon is complex as it is affected by a number of variables such as pH, dissolved oxygen content, temperature, quality of flowing fluid, quality of oxide layer on inner surface of the pipe, chemical composition of the steel pipe and particle impact angle [2]. Many researchers have made attempts to develop models for estimation of erosion-corrosion rate and to formulate service life models for piping components subjected to erosion-corrosion degradation mechanism. Stack and co-workers developed a mathematical model for estimating erosion-corrosion in mild steel pipes carrying aqueous solution containing alumina particles based on detailed laboratory studies [2]. They assumed the erosion-corrosion process to be purely additive, i.e., sum of erosion and corrosion effects. The model is admittedly applicable for low particle impact angles (impact angles <4^{o}), low flow velocities (flow velocities <2 m/s), constant temperature and constant pH (pH = 9.0) of flowing fluid. Abdulsalam and Stanley [3] developed a steady state model to account for the steady hydrogen flux through metal and has established that erosion-corrosion is dependent on the kinetic rate of metal oxide film dissolution at lower temperatures and on mass transfer limited rate at higher temperatures. Ting and Ma [1] developed an erosion-corrosion model based on phenomenological considerations and statistical data of pipe wall thickness obtained from Taiwan PWR nuclear power plants for different piping components subject to various operating conditions. In the present investigation, the model proposed by Abdulsalam and Stanley [3] is used for estimating the EC rate. All the three models discussed above are deterministic.

2.1 Need for Stochastic Modeling of Erosion-Corrosion

Due to the uncertainties in material properties of steel and the variations in exposure conditions, the degraded state of the piping component subjected to erosion-corrosion will be a random variable at any given time. Also, the degraded state of the piping component changes with time. Thus, the evolution of the degradation in the piping component has to be modeled as a stochastic process. Markov Chains (MC) are found to be a useful tool for stochastic modeling of condition state evolution of degrading systems [4, 5]. A homogeneous MC model for assessment of piping components against erosion-corrosion is presented in the next section.

3 Markov Chain Modelling

The degraded state of the piping component (hereafter referred to as system) subjected to erosion-corrosion will be a random variable. Also, the degraded state of the system changes with time. Thus, the evolution of the degradation in the system has to be modelled as a stochastic process. Markov chain (MC) models are the simplest stochastic models that are extensively applied in engineering [4–7]. In a Markov Chain model, both the state space and the index space can be discrete.

In the case of erosion-corrosion, the state space is the degradation state (z) of the system represented by the ratio of the loss in wall thickness due to erosion-corrosion (l) to the original wall thickness of the pipe (t), and the index space is the time ($T = \{T_1, T_2, \ldots, T_n\}$). The loss in wall thickness is given by

$$l = A \times WR \times \text{Age (in years)} \tag{1}$$

where WR is the rate of erosion-corrosion per year and A is the modelling error. The value of WR can be determined using a suitable erosion-corrosion model.

The probabilistic evolution of the process, in general, can be described by the transition probabilities,

$$TP = \mathrm{P}\{z(T_i) = i|z(T_{i-1}) = i-1, z(T_{i-1}) = i-2, \ldots, z(T_1) = 1\} \tag{2}$$

In this study, the probabilistic evolution of degradation is obtained by making the following assumptions.

i. The stochastic process can be described as a one-step memory process. This implies that the process is Markov and present state of the system can be completely determined by its immediate past state. This assumption is justified since degradation at *i*th time (i.e. at time step t_i) more or less depends on degradation at (*i-1*)th time (i.e. at time step t_{i-1}).
ii. The stochastic process has a discrete, finite state space {1, 2, …, m}, and, a discrete index space {1, 2, …, n}, where index 1 is interpreted as time step = T_1, index 2 is interpreted as time step = T_2, and so on. Since loss in wall thickness increases with age, the system can make transitions from a given state to the higher states only.

Using these assumptions, transition probability for the system is given by,

$$p_{ij}(T_k, T_{k+1}) = \mathrm{P}\{\varepsilon(T_{k+1}) = j | \varepsilon(T_k) = i\}; \quad 1 \leq i \leq m, \quad i \leq j \leq m, \\ 1 \leq k \leq n-1 \tag{3}$$

The probabilistic evolution of the system is given by the transition probability matrix (TPM),

$$P(T_k, T_{k+1}) = \left[p_{ij}(T_k, T_{k+1})\right]_{1 \leq i \leq m, i \leq j \leq m}, for \quad 1 \leq k \leq n-1 \tag{4}$$

Since the system can make transitions from a given state to the higher states only, the TPM will be an upper triangular matrix. Since the state space considered is such that the states are mutually exclusive and collectively exhaustive,

$$\sum_{j=1}^{m} p_{ij}(T_k, T_{k+1}) = 1, for \quad 1 \leq i \leq m \tag{5}$$

3.1 Determination of k-Step TPM

The probabilistic description of the state of degradation after *k*-time steps is given by (Chapman Kolmogorov equation),

$$P(T_1, T_k) = P(T_1, T_2) \times P(T_2, T_3) \times P(T_3, T_4) \times \ldots \times P(T_{k-1}, T_k) \tag{6}$$

Since a homogeneous Markov Chain is considered in this study, $P(T_i, T_{i+1}) = P(T_{i-1}, T_i)$. Hence, the *k*-step TPM is given by $P(T_1, T_k) = P^k(T_1, T_2)$. The unconditional probability vector of the state of degradation, after *k*-time steps can be determined from,

$$\left(P^U(T_1, T_k)\right)_{1\times m} = (P(0))_{1\times m} \times \left[P^k(T_1, T_2)\right]_{m\times m} \tag{7}$$

where $(P(0))_{1\times m}$ is the vector representing the probabilities of initial states of the system. For a system whose evolution is defined by a homogeneous MC, the state of the system at any future time can be determined using the one-step TPM, once the initial state is known.

3.2 Determination of Elements of TPM

A typical element of 1-step TPM (Eq. 3), can be written as,

$$p_{ij}(T_k, T_{k+1}) = \frac{P\{z(T_{k+1}) = j \cap z(T_k) = i\}}{P\{z(T_k) = i\}} \tag{8}$$

which gives the probability of degradation state of the system being '*j*' at time T_{k+1} given that the degradation state was '*i*' at time T_k. Computation of these probabilities requires information regarding joint probability density function (jpdf) of degradation state at any two successive time steps, (T_k, T_{k+1}) and pdf of degradation state at time step, T_k. Since it is difficult to generate this information from test data, in the present investigation, it is assumed that degradation states at successive time steps follow bivariate normal distributions and at any time step, degradation state follows a normal distribution. This is because when the mean and variance are the only information available with respect to the degradation state of the system at any time step, the maximum entropy distribution is the normal distribution [8]. Hence, it is assumed that the state of degradation at any load step follows normal distribution. Knowing the jpdf and pdf, and using Eq. (7), the elements of TPM can be computed. A typical element of the conditional 1-step TPM is given by

$$p_{ij}(T_k, T_{k+1}) = \frac{\int_{z_{i-1}}^{z_i} \int_{z_{j-1}}^{z_j} f_{k,k+1}(z_k, z_{k+1}) dz_k dz_{k+1}}{\int_{z_{i-1}}^{z_i} f_k(z_k) dz_k} \tag{9}$$

where $f_{k,k+1}(z_k, z_{k+1})$ is the bivariate normal distribution with correlation coefficient $\rho_{k,k+1}$ and $f_k(z_k)$ is the univariate normal distribution.

The step-by-step procedure for MC modelling of the degradation in piping component is given below.

1. Divide the state space into mutually exclusive and collectively exhaustive event sets.
2. Divide the index space into discrete intervals.

3. Compute the mean and standard deviation of the degradation state of the system under the considered degradation mechanism at two successive points in the index space.
4. Using the values of mean and standard deviation computed in step 3 and using a suitable correlation coefficient ($\rho_{k,k+1}$), formulate the one-step TPM, P, using Eq. (9).
5. Determine unconditional probability vector of the state of degradation of system after k-time steps using Eq. (7).

A software, called RISCMarkov, is developed at CSIR-SERC for reliability analysis of power plant piping components (Fig. 1) [9]. The software can be used for MC modeling of piping components against erosion-corrosion, thermal fatigue, vibration fatigue and stress corrosion cracking.

The state space, given by $z = l/t$, is between 0 and 1, and is divided into a finite number of discrete states as defined by the user (default number of divisions is taken as 20 in the software). The index space is discretised into one year intervals. Depending upon the operating conditions and the inputs available, the software has four options (Fig. 2) for determining the rate of erosion-corrosion (*WR*). The values of mean and standard deviation of z at two successive years is obtained using first order approximation, and the n-step TPM is computed using the step-by-step procedure given above. The state probabilities corresponding to four states defined as *success* (no detectable damage; l/t < 0.125), *flaw* (detectable flaw; 0.125 ≤ l/t < 0.45), *leak* (0.45 ≤ l/t < 0.80) and *rupture* (l/t ≥ 0.80) are determined, using the aggregation procedure given in Balaji Rao and Appa Rao [8].

RISCMarkov

SOFTWARE FOR RELIABILITY ANALYSIS OF PIPING COMPONENTS USING MARKOV CHAIN APPROACH

DEVELOPED AT SERC, CHENNAI

Fig. 1 RISCMarkov

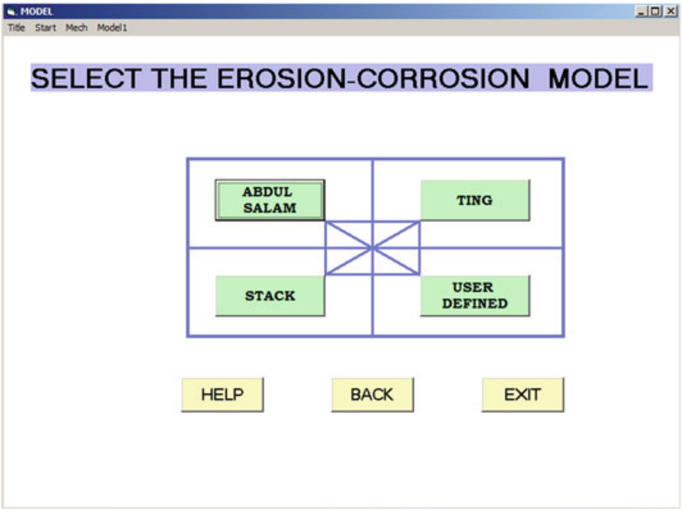

SELECT THE EROSION-CORROSION MODEL

Fig. 2 Options for the erosion-corrosion model in RISCMarkov

4 Example

An outlet feeder pipe of a PHWR is considered. The feeder pipe is made of carbon steel A106GrB. In the present study, the model proposed by Abdulsalam and Stanley [3] is used for determining the rate of erosion-corrosion and hence the loss in wall thickness at different times. The diameter of the pipe is 70 mm and thickness is 6.5 mm. The flow velocity is 1500 cm/s, the pH is 10.2, and the temperature is 553 K. The kinematic viscosity is taken as 0.0179 cm^2/s. The plant life is taken as 40 years. The random variables considered, along with their mean and standard deviation values are given in Table 1. The vector representing the unconditional probabilities of the initial states of the system $((P(0))_{1\times m})$ is taken as $\{1, 0, 0, \ldots, 0\}_{1\times 20}$, since the loss in wall thickness due to erosion-corrosion is zero at the beginning. To study the effect of correlation coefficient $\rho_{k,k+1}$ on the state probabilities, three values of $\rho_{k,k+1}$, namely, 0.0, 0.5 and 0.99, are considered.

Table 1 Random variables considered

Variable	Mean	COV
Pipe diameter (cm)	7.0	0.0174
Pipewall thickness, t (mm)	6.5	0.059
Temperature (Kelvin)	553	0.009
pH	10.2	0.07
Flow velocity (cm/s)	1500	0.005
Model error	1.0	0.01

5 Results and Discussion

Using the proposed MC model, the variation in unconditional state probabilities with time during the life of the plant (40 years) are determined, and are shown in Figs. 3, 4 and 5 for the different values of $\rho_{k,k+1}$ considered. From these figures, it is noted that as $\rho_{k,k+1}$ increases, the relative time spent in the intermediate degradation states (namely, flaw and leak) reduces. This suggests that, for lower values of $\rho_{k,k+1}$, the decision making regarding in-service inspection is governed by probabilities of the piping component being in flaw and leak states. The variation in probability of failure against rupture with time for the three values of $\rho_{k,k+1}$ considered are shown in Fig. 6. Since the values of failure probability against rupture are small for $\rho_{k,k+1}$ = 0 and 0.5, the failure probabilities are shown in logarithmic scale in Fig. 6. From Fig. 6, it is noted that as $\rho_{k,k+1}$ decreases, the probability of finding the system in rupture state also decreases. This suggests that if the dependence is not considered in modeling the evolution of degradation of the system, the probability of failure values obtained can be unconservative. Integrating the values of state probabilities at

Fig. 3 Variation in unconditional probabilities of states with time for $\rho_{k,k+1}$ = 0

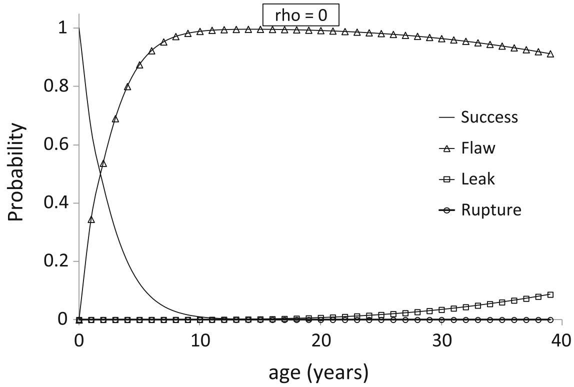
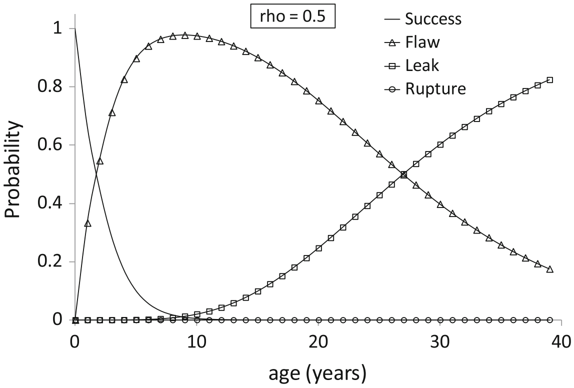


Fig. 4 Variation in unconditional probabilities of states with time for $\rho_{k,k+1}$ = 0.5



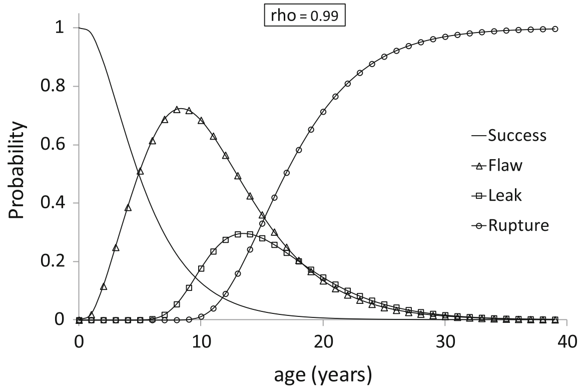


Fig. 5 Variation in unconditional probabilities of states with time for $\rho_{k,k+1} = 0.99$

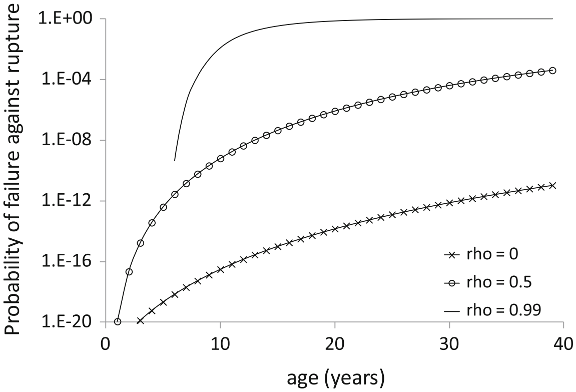


Fig. 6 Probabilities of failure against rupture

different times obtained using the MC model with the consequences associated with the piping component being in different degradation states, will be useful for risk informed in-service inspection of these components.

6 Summary

A homogeneous MC model for probabilistic failure assessment of piping components against erosion-corrosion is presented. The MC model is incorporated in the software RISCMarkov, developed at CSIR-SERC for reliability analysis of power plant piping components. Using the MC model, the variations in failure probabilities against rupture with time for a power plant piping component are determined. The unconditional state probabilities at different times obtained using the MC model can be integrated with the consequences associated with the piping component being in different degradation states for risk informed in-service inspection of these components.

Acknowledgments The paper is being published with the kind permission of Director, CSIR-SERC, Chennai. The authors are thankful to Board of Research in Nuclear Sciences for partially funding the work (Sanction no. 2000/36/13/BRNS) which was carried out during 2001-2004.

References

1. Ting K, Ma YP (1999) The evaluation of erosion/corrosion problems of carbon steel piping in Taiwan PWR nuclear power plants. Nucl Eng Des 191(2):231–243
2. Stack M, Corlett N, Turgoose S (1999) Some recent advances in the development of theoretical approaches for the construction of erosion-corrosion maps. Wear 233–235:535–541
3. Abdulsalam M, Stanley J (1992) Steady-state model for erosion-corrosion of feedwater piping. Corrosion 48(7):587–593
4. Ang AHS, Tang WH (1984) Probability concepts in engineering planning and design vol ii, Decision risk and reliability. Wiley, New York
5. Fleming KN (2004) Markov models for evaluating risk-informed in-service inspection strategies for nuclear power plant piping systems. Reliab Eng Syst Saf 83(1):27–45
6. Balaji Rao K, Appa Rao TVSR (2004) Stochastic modelling of crackwidth in reinforced concrete beams subjected to fatigue loading. Eng Struct 26:659–667
7. Anoop MB, Balaji Rao K, Lakshmanan N, Raghuprasad BK (2012) Markov chain modeling of evolution of strains in reinforced concrete flexural beams. Mater de Construcción 61(307):443–453
8. Kapur JN (1993) Maximum Entropy Models in Science and Engineering. Wiley Eastern Limited, New Delhi
9. Balaji Rao K, Anoop MB, Lakshmanan N, Gopika V, Saraf RK, Kushwaha HS (2004) A methodology for risk informed in-service inspection for safety related systems—Final Report, Report No. SS-GAP01241-RR-04–3

Turnout Degradation Modelling Using New Inspection Technologies: A Literature Review

Niloofar Minbashi, Morteza Bagheri, Amir Golroo, Iman Arasteh Khouy and Alireza Ahmadi

Abstract Turnouts are of the most critical components of railway track which are prone to high static and dynamic forces leading to more intense degradation. They require more inspection than other parts of a railway track as they are potential safety hazards. As a result, turnout degradation processes are crucial to be understood by infrastructure manager to plan for their maintenance and renewal in advance. Two approaches have been introduced in the literature to achieve a thorough understanding of degradation processes in turnouts. The first one acts to develop degradation models based on influential parameters and historical data and then to predict degradation processes in the future; while the second one tries to improve inspection through using new concepts and technologies leading turnout condition data to be better captured over time. The purpose of this paper is to review all available resources regarding these two approaches and provide a guide for further research into turnout studies.

Keywords Railway turnouts · Degradation processes · Inspection technologies · Degradation models · Railway transportation

N. Minbashi · M. Bagheri (✉)
School of Railway Engineering, Iran University of Science and Technology, Tehran, Iran
e-mail: morteza.bagheri@iust.ac.ir

N. Minbashi
e-mail: minbashi@rail.iust.ac.ir

A. Golroo
Department of Civil and Environmental Engineering, Amirkabir University of Technology, Tehran, Iran
e-mail: agolroo@aut.ac.ir

I. Arasteh Khouy
Luleå Railway Research Centre (JVTC), Luleå University of Technology, Luleå, Sweden
e-mail: iman.arastehkhouy@ltu.se

A. Ahmadi
Department of Civil Environmental and Natural Resources Engineering,
Luleå University of Technology, Luleå, Sweden
e-mail: alireza.ahmadi@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_5

1 Introduction

All around the world, railway industry is planning for higher transit speeds and extended capacity for freight transportation. This means that railway assets are more prone to degradation than before, so that their reliability should be enhanced as they are to meet increasing demands of the future.

Turnouts are among the most crucial assets in a railway system, as they provide flexibility and punctuality to the railway network, particularly when a disruption occurs; they allow trains to use routes other than usual to ensure reliable services for trains and passengers. Therefore, a substantial portion of railway network budget is spent on their annual maintenance and required renewals. For example, maintenance cost of turnouts comprises a minimum of 13 % of the total maintenance cost of the railway network in Sweden [1]. That is even more in Switzerland, where 25 % of the railway maintenance and renewal budget is spent on turnouts [2]. However, In the United States, turnouts are identified as a major cost area with an annual budget being estimated to be 10 times more than the amount spent on conventional track [3]. Turnouts have a distinct structure within railway assets, hence they need to be taken care of more cautiously. The distinctive structure of turnouts comes from the following: turnouts have special components, such as switch tongs, frogs and slide plates, which are prone to high vertical and lateral dynamic forces because of their particular geometry leading to a considerable amount of deterioration [4]. Another aspect regarding turnout structure is that turnouts can be considered as a mechanical system as they have moving parts, meaning that more inspections and maintenance actions are needed to assure their reliability; last, but not the least, is that turnouts are considered to be potential safety hazards. In the United States, approximately 10 % of the derailments on yard and siding tracks have been caused by turnout defects coming from the heavy use of turnouts on these types of tracks leading more wear and deterioration to be imposed on them [5].

Degradation processes in turnouts must be monitored in order to plan for maintenance activities ahead. This is possible if the conditions of turnout and its components are available over time implying the importance of high quality data availability for analysing turnout condition. For turnouts, unavailability of reliable data has been a crucial problem, as no preventive maintenance can be planned without a reliable dataset. Prediction of maintenance and renewal requirements of turnouts, like any other railway asset, is possible once degradation processes are known and predictable. So far few studies have been carried out to model degradation processes of turnouts to improve their maintenance planning [6]. However, there are two approaches for understanding the way turnouts degrade: the first approach advances via developing degradation models for prediction of maintenance and renewal requirements of turnouts based on historical data [2]. The second approach works through inspection and documentation of inspection tasks which has become possible recently by developing new technologies for data collection and defect analysis [7]. The aim of using new inspection technologies is to document turnout inspections over time which enables the infrastructure manager to analyse degradation trends when making decision for each turnout individually, based on its condition.

The purpose of this paper is to review two approaches toward understanding turnout degradation: degradation processes modelling using historical data and new inspection technologies. This paper provides a thorough scheme of all the attempts carried out to reflect the degradation processes of turnouts. The first section addresses the approach of using historical data to model degradation. In the second section, new inspection technologies are reviewed comprising new devices for semi-automated data collection as well as new technologies for automated data collection, while the conclusions are presented in the final section.

2 Degradation Modelling

If the degradation of a turnout was identified to have reached its critical level, then safe operation of trains may not be guaranteed anymore. Therefore, a maintenance or renewal action is needed to avoid any hazardous situation to be faced. Degradation can be predicted enabling the infrastructure manager to take decision for maintenance or renewal of a component or a series of components. This prediction may be based on a maintenance index defined by measurement of components during inspection phase, as will be described in the next section, or by implementing models based on historical data. In the case of turnouts, it is hard to implement degradation models based on historical data because of unavailability of data, so as only two cases have focused on modelling of degradation processes of turnouts, as it can be seen in the following paragraphs.

Zwanenburg [2] implemented a model for degradation processes of turnouts. This model is the first model available so far in the literature on degradation of turnouts. The purpose of the model is to determine maintenance and renewal requirements of turnouts for mid-term planning over a period of 10 years. Degradation processes stand for degradation and wear of the turnouts; the former is a reduction in the quality of track geometry, while the latter is a reduction in that of components.

Parameter selection has been based on three categories: (1) train (axle load, total tonnage), (2) track (soil quality, maintenance and component renewal policy), and (3) operation (whether trains are mainly in facing or trailing direction, train speed).

The model has been based on maintenance and renewal data from the Swiss Federal Railways. Maintenance tasks included in the model based on the available data were: (1) tamping (geometry correction), (2) welding on the frog, and (3) grinding of metal parts. The following options were available for replacement of the components: (1) complete switch or crossings replacement, (2) switch rail with accompanying stock rail, (3) frog, and (4) check rail.

In this model, degradation and wear have been described by developing a reliability model where reliability is defined as the probability of a turnout or its components to function properly longer than a specified period of time.

The model resulted in approving parameters assumed to be influential on the degradation of standard turnouts, such as: (1) the actual train load, (2) lower soil quality which reduces life expectancy of a switch, (3) smaller switch angle which is generally associated with a longer life, (4) higher axle loads (more freight trains) leading to more wear, and (5) train speed. At the end, taking a sample from another period of time, geographical area or railway network has been recommended because the proposed model failed to be successful for the Swiss data.

Zwanenburg [2] tried to model geometrical degradation and wear of turnouts. However, his model is just a mean to reveal the importance of the parameters. Arasteh Khouy et al. [8] focused on geometrical degradation of turnouts and attempted to analyze vertical geometry degradation using longitudinal level measurements over a four-year period. The reason for this comes from the fact that geometrical condition of the track can trigger degradation of other track components and hence is used to assign entire range of track maintenance operations [9].

Geometrical degradation at the crossing point of turnouts has been analyzed by two approaches. The first approach considers two parameters for analyzing geometrical degradation, namely, the absolute residual area (AR_a) defined as "*the absolute value of the area obtained from the differences in the longitudinal level values between two adjusted measurements at the crossing point*" and maximum settlement (Smax). defined as "*the difference between the value of longitudinal level at the crossing point and the value obtained from the vertical line passing through the crossing point line connecting the positive peaks before and after the crossing point*". The absolute residual area (AR_a) can indicate the trend of track settlement due to accumulated loading over a certain period of time. In the analyzing phase, the first measurement is taken as the reference point to which subsequent measurements are compared, meaning that the longitudinal level of the reference point is assumed to be constant. This results in the estimation of the relative geometrical degradation rather than the current one. Therefore, analysis of geometrical parameters of the crossing points, such as the slope of measurement line at 1 m before and after the crossing point, has been used in the second approach. The trends of these parameters as a function of time have been analyzed. The results of the second approach reveal that crossing position settlement has a limit after which the crossing cannot settle anymore and the faults would be transferred to the next wave in the crossing neighborhood.

These two approaches using different parameters reveal that turnouts should be regarded individually since they are associated with different degradation rates. The difference in degradation rates comes from other factors which were not considered in the study, such as traffic, subgrade quality, age of the asset, maintenance strategy and the environment. Even though, Arasteh Khouy et al. [8] did not introduce a geometrical degradation model for turnouts, but their work provides a reliable knowledge for better understanding of the turnout settlement. This knowledge can be incorporated into a LCC model in terms of specifying maintenance intervention limits considering the cost effectiveness.

3 Data Collection

Safe rail operations are guaranteed by periodic inspection of railway infrastructure. However, this is a hard task to accomplish as tight train scheduling doesn't allow much flexibility within inspection operations. Therefore, new technologies are introduced to improve inspection task especially for turnouts which are inspected manually with their inspection being a labour-intensive duty for infrastructure manager. In this section new technologies for better manual inspection of turnouts are introduced.

3.1 Semi-automated Data Collection

Turnout inspection in its traditional form where a paper was used for checking and recording the turnout condition, was upgraded by introduction of palmtop computer systems called SwitchInspect by ZETA-TECH. Zarembski [10] introduced this Personal Digital Assistant (PDA) to be used in turnout inspection. The data gathered by this handheld computer can be uploaded to a database enabling the infrastructure manager to prioritize maintenance activities and schedule them. The proper inspection data help us make an overall rating of turnout condition by defining turnout indices that cintribute to prioritization of turnout maintenance tasks or turnout renewal and also to a safety assessment of the turnout condition. More importantly, the definition of indices lead to a thorough understanding of degradation process of turnout enabling the infrastructure manager to plan for maintenance and renewal of turnouts before a breakdown.

There are seven component areas with their related components in SwitchInspect system: 1. geometry, 2. switch stand, 3. switch point area, 4. closure area, 5. frog area, 6. ballast and 7. ties.

This system has two phases. The first phase is data collection phase where condition indices are determined and then employed to plan maintenance activities ahead.

In data collection phase, the inspection measurements are recorded and components are ranked based on their relative importance from view point of operation and maintenance. All the possible failure modes with their severity are also included in the device for each component. Importance of each turnout and traffic density are included as well.

The second phase comprises the calculation of the condition and maintenance indices. Two kinds of indices are calculated based on collected data: an overall Turnout Condition Index and Maintenance Sub-Indices (MSI) for each turnout and key maintenance areas which consist of: 1. tamping, 2. gaging, 3. grinding, 4. welding, 5. tie replacement, and 6. switch stand replacement. Maintenance Sub-Indices Calculation is based on a series of numerical ratings related to each inspection item, its mode of degradation/failure with its severity taking the importance of the turnout itself into account [10]. The result of the second phase is a set of Maintenance Sub-Indices (MSI) which can be used together with or separately from Turnout Condition Index (TCI) to prioritize scheduled turnout maintenance and also to determine the necessity of renewal of a complete turnout or its components. Maintenance Sub-Indices are particularly useful as they can prioritize specific maintenance activities. Prioritization of turnout maintenance and renewal operations is an important issue because of the large number of turnouts in the network and their degradation rate which is crucial to high-density traffic lines.

SwitchInspect allows a systematic documentation of turnout inspection to be performed manually. Its main feature is that the whole condition of turnout can be documented, so as degradation of the turnout can be traced over time. However, the inspection process is still dependent on the inspector and prone to human error. There are other devices identifying the wear condition of turnout, such as MiniProf Switch which is an add-on to MiniProf Rail. The MiniProf Rail provides instant information on metal removal and grinding stone tilt. The MiniProf Switch is the extended MiniProf Rail which is able to measure special profiles of switches. It is magnetically attached to the rail and is able to use a telescopic rod for reference to the opposite rail [11].

Each rail in a switch or crossing is measured individually, then the positional data (reference profiles) from the Switch add-on are combined to these measurements automatically. The measurement containing all profiles placed relative to each other allowing measures to be conducted between them is the result of this device [11]. It can also be used as a PDA because of its embedded battery. USB connection to the instrument is the feature that makes it possible to use the measurement data for other modelling purposes. The MiniProf Switch instrument is a useful device for switch inspection, however, it can be used for only one maintenance area within turnout inspection which is grinding [11].

A laser based trolley for switch and crossing inspection has been introduced in [12]. The aim of this system is the efficient profile inspection of switches and crossings and advising on their maintenance by means of welding and grinding. The proposed solution is a lightweight trolley of an estimated weight of 15 kg which can be manually pushed over the switch. The relative position of the trolley can be monitored by a tachometer. Profile data acquisition from the parts to be inspected is done through separate 2D transversal slices or as a complete 3D model. For profile acquisition, two line lasers will be used to scan the rail's profile. These lasers are selected based on such features as their width of the scan, number of samples per line of scan, precision, sampling frequency, size, power consumption, robustness, price and overall quality. The utilized laser for this application is a scanCONTROL 2700–100 [13] produced by Micro Epsilon which enables to scan half set of one switch at a time (one stock rail and one switch rail). The crossing is inspected by pushing the trolley two times, once on a straight path and one other time on a diverging path. A 3D profile scan of the crossing is required by inspection standards and needs to be identified for this system; this is introduced as a further research. The required power for the trolley will be provided by rechargeable batteries; it comes with a tablet computer and a software written in LabView programming language. Ethernet port will provide the communication between the lasers and the tablet as it is easier to work on an interface provided by a tablet computer. The inspection of different parts of the switch is possible for the user and also he/she is advised on how the profile should be rectified. Furthermore, he/she will be notified about the accordance of the profile with the Network Rail standards. The fact that this system is used manually is defined as its main advantage because it provides a good flexibility and implements the criteria defined by the inventors [12].

Jönsson et al. [14] focused on turnout geometry inspection of switch blade and frog section as track geometry quality can cause other components degradation [13]. They introduced a new method to measure the vertical position of unstressed track geometry relative to the main track over time. The relevant measurements can be used to develop degradation models which will then be used in life cycle cost (LCC) models. The aim of inspecting geometrical shape of turnout is to ensure its proper function and safe passage of the trains, since in case of improper geometrical shape, necessary maintenance actions would be costly. For the measurement of longitudinal level (vertical direction) under unstressed conditions, four measuring equipment have been evaluated based on their accuracy in terms of technical specifications, repeatability and practical accuracy, setup time, range of measurements and number of users required to carry out the measurements. The levelling instrument has been selected for measuring geometric irregularities of the track. During three measurements in different periods, 13 turnouts with different ambient and operating conditions are selected. The overall results showed that the dependability of a railway infrastructure is affected by the seasons, meaning that climate conditions, such as frost or ground water levels, can change geometrical shape of turnouts [14]. The Million Gross Tons (MGT) can also increase vertical track geometry elevation/decline of turnouts towards their mid-section. The largest vertical track geometry decline was related to turnouts with a radius on the through route. The turnouts had a relatively higher vertical positions in comparison with the main track which may come from this fact that differences in history of tamping frequency as well as stiffness and load carrying area.

3.2 Automated Data Collection

Periodic track inspection is a requirement of applicable standards for individual railroad track maintenance to guarantee a safe and efficient operation. Even though introducing new devices, such as SwitchInspect and MiniProf Switch, has simplified the inspection process for inspectors but it is still a labour-intensive, time-consuming and slow task. Therefore, new technologies are required to automatically inspect railway turnout resulting in the elimination of human error. In recent years, new technologies have been developed for automatic inspection of turnouts.

The Netherlander company, Eurorailscout, developed a switch inspection system in 2005 which was later equipped with a profile measurement system for wear detection in 2009. The whole system is called “Switch Inspection and Measurement”

(SIM). This system is made up of a wagon which is pulled and pushed on track for turnout inspection. The inspection system is made up of two panorama cameras which are placed toward the front and the rear of the wagon to capture the perspective of the running inspection [15]. Four line-scan cameras provide recordings of the outer and inner sides of the rails, including the inner parts of the switches. Two line-scan colour cameras capture top view of the rail, while two black and white cameras pride detailed view from the rail head to find cracks and other anomalies. The system also uses an intelligent software for analysis and examination of potential defects which leads to automatic identification of defects such as missing fastening devices, crumbling of concrete rail sleepers, cracks in the concrete rail sleepers, ballast deficit and ballast surplus [15]. The measurement system uses a well-known triangulation method which means that the travelling route is determined by a spot laser whose reflection is monitored by the camera. Then, the distance between the camera and object can be calculated from such reflection data. Track profile is once scanned per 20 mm, at a measurement speed of 40 km/h. The profile measurements can then be presented to the client in any desired reporting format: txt, csv, xls, pdf, xml, etc. The horizontal and vertical wear can be calculated from measured profiles as well. The inspection may involve different cross-section measurements at the tongue, the track or the frog. All the data and reports will be kept in a data bank. The measurement history preservation in this data bank can contribute to trend analyses. SIM is able to measure 3D geometry of tracks and switches by an inertial measurement system optimised for very short and median-length waves. The delivered parameters are track width, shift, height, transverse gradient in accordance with EN 13848. The inertial system is able to attain high accuracy, both at low speeds and also with very short waves at high speeds [15].

Molina et al. [16] brought the concept of machine vision for turnout inspection. This concept had been previously developed for other rail infrastructures such as: (1) rail surface defects [17, 18], (2) rail wear [19], (3) tie condition [20–22], (4) ballast [23], (5) fastening systems [24, 25], and (6) general track structure inspection [26, 27]. Using machine vision techniques in the course of railway inspection systems has been beneficial as some experimental tests showed, for example, that accuracies have been greater than 80 % with measurement speeds of up 320 km/h in many cases [16]. Molina et al. [16] have reviewed previous researches into machine vision applications in railway infrastructure inspection and mentioned that no previous work has been dedicated to turnout inspection using machine vision. As a result, they decided to develop an algorithm for detecting most critical components of a turnout. They analysed turnout-related derailment data from 1998 to 2009 using FRA Accident Database identifying a rank-ordered turnout components/defects selection to be inspected using machine vision: (1) worn or broken switch point, (2) other frog, switch, and track appliance defects, (3) worn or broken turnout frog, (4) broken or defective switch connecting or operating rod, (5) gap between switch point and stock rail, (6) missing bolts and cotter pins. As the inspection of missing bolts and cotter pins was the first task done by means of visual inspection, they decided to select it as their first inspection priority using machine vision. Developing machine vision system is based on collecting video and images from track components. This is critical as the component must not be shown just in its functional position, but it must also be distinguished from the background objects and provide necessary measurements by correct orientation. They selected two camera views for inspection, the lateral view and over-the-rail-view. These views were suitable for detecting tie plates, anchors and spikes. A Video Track Cart (VTC) was designed for collecting continuous video shots of sections from a low-density track for experimental data acquisition [16]. They captured videos from tangent and different turnouts under different conditions in terms of natural lighting, levels of vegetation, ballast types and levels of ballast fouling. They developed an algorithm for detection of spike, anchor and tie and recognition of defects which goes through a coarse-to-fine approach for object detection. It detects the track components with predictable location, such as rail, before locating the objects with high appearance variability, such as spike heads and anchors. Local features, such as edges and texture information are also incorporated into the model in order to increase the robustness to changing environmental conditions. The introduced system has shown good reliability for component inspection using machine vision, nevertheless, the algorithms need to be refined to improve the reliability of spike and anchor detection.

Afshari et al. [28] identified bolted joint in turnouts as one of the most hazardous components causing accidents. Therefore, they introduced an effective method for health monitoring of bolted joints in railroad switches. They claimed that early detection of loosed bolted joints using a full-automatic mechanism to inspect the switches' mechanical condition would eliminate the need for frequent visual inspections. They have applied piezoelectric transducers and impedance-based structural health monitoring techniques for monitoring the loosening of bolted joints in a full-scale railroad switch. The results have shown that a quarter turn of a bolt could be clearly detected by measuring the electrical impedance of a PZT patch at the bolted connection. The accuracy of this system is as high as 25 ft-lbs when the bolt assembly is loosened, which corresponds to merely one tenth of a bolt turn. The experimental results showed that each PZT sensor/actuator attached to a bolts' nuts is more sensitive to its corresponding bolt rather than neighboring ones, meaning that the proposed damage detection is able to isolate the loosened bolt from the others. After detection of the loosed bolted joints, the loosened bolts are retightened using a shape memory alloy (SMA) washer as the actuator. The retightening comes from a self-healing concept which is integrated into the impedance-base structural health monitoring technique (SHM) making the inspection bolted joints fully automated leading to eliminate the need for frequent visual inspections.

Schoone [29] invented a contactless system for capturing the profile of a rail of a turnout in order to determine its conditions in terms of wear and deformation. Wear and deformation of turnout is corrected via replacement or grinding and is then evaluated by periodic inspections. Periodic inspection of turnout rail can be realized using contact or non-contact sensors, e.g. laser measurements known as triangulation. The proposed system uses a laser device to project a light beam onto an area of the rail facing the opposite rail before the reflected light from that area of the rail is recorded by an imaging device. In this system, the location of the point blade of the turnout is detected as the reference point and the apparatus has the capability to be mounted on a passing train. The aim of this system is (1) to evaluate the wear and deformation of the turnout through parameters measurement of components such as rail, without any contact, (2) to acquire the actual profile by combining information from the components, (3) to accurately locate measurement point relative to the reference point, (4) to obtain results from measurements in real time, and (5) to conduct measurement at 10 or 20 km/h of the measuring train velocities.

Zarembski et al. [30] introduced the very new concept of using an automated inspection vehicle for switch inspection. The idea has come from the fact that manual process of inspection did not change even with using PDAs and it was still dependent on the corresponding inspector. Automated Switch Inspection Vehicle (ASIV) is a new technology dedicated to turnout inspection. This vehicle is able to test turnouts at a high degree of measurement accuracy and frequency generating an appropriate level of information on turnout condition in order to monitor them from safety and maintenance management point of view.

ASIV contains a profile measurement system which measures the switch and frog profiles along with a newly developed state of the art switch analysis software (SwitchWear) for analysing the measured profiles.

ASIV is able to measure switch point, frog and stock rail profiles, wear and important geometry parameters. It can identify derailment problems, damage related to switch points, stock rails and frogs as well as, wide gaps on closed switches. By generating data from rail condition, turnout degradation can be monitored over time and, more importantly, one can identify safety hazards and conditions violating FRA safety rules and also railroad maintenance conditions.

In a technical sense, ASIV can be defined in this way: ASIV uses a high-sampling-rate profile acquisition for image acquisition at one inch intervals at 8 mph ($\sim$12.9 km/h) speed on each rail. For instance, for a specific turnout of 170 ft ($\sim$52 m) in length, more than 8000 rail profiles will be acquired with the straight and diverging legs of the turnout, and the whole turnout will be measured in less than five minutes.

After acquiring rail profile data, 3-D composite images of turnout and its key components are developed. These composite images can be used for the analysis of key maintenance and safety parameters (Table 1) in comparison with a specific standard which can be FRA Track Safety Standards or the railroad-specific maintenance standard. When the relative deviation from these standards is labelled as Red defect, it means that a safety standard violation has been occurred, while if it is labelled as Yellow, a maintenance standard has been exceeded without violating safety standards. Red and Yellow defects data will be used for calculation of Turnout Maintenance Priority Index for each turnout determining its overall condition and priority in terms of maintenance or renewal. Using this index when evaluating overall conditions of all turnouts, will lead to identify high priority turnouts with very bad conditions within mainline or rail yards.

Identification of defects leading to component failure and derailment by ASIV contributes to derailment risk mitigation which is realized by analysing the interactions between wheel and rail. SwitchWear is able to analyse various turnout

Table 1 Summary of potential measurements in the switch area by the ASIV [29]

Rail type	Measurement
Stock rail opposite a switch rail	Vertical wear
	Gage side wear
	Gage face angle
	Gage corner radius
Switch rail	Gage face angle
	Breaking or chipping
	Gage corner radius
Stock + switch rails	Vertical height difference
	Lateral gap width
	Wheel contact point through switch point
Closure rails	Vertical wear
	Gage side wear
	Gage face angle
Guard rail	Guard flangeway gap width
	Relative height of guard rail
Frog nose and wing rail	Relative height of nose and wing rail
	Wear/Batter on Wing Rail
	Batter/damage to frog
	Flangeway depth
	Flangeway width
	Surface damage: Batter, chipping
	Wheel contact through frog
	Wing rail profile (within field of view)

surface geometry and multiple wheel profiles. ASIV develops a Turnout Derailment Risk Index which addresses potential risk of derailment [29].

In conclusion, ASIV allows for monitoring deterioration of turnout by ongoing measurements. Therefore, maintenance approach will become proactive rather than reactive. ZETA-TECH is trying to upgrade ASIV by enabling the inspection of other components of turnout including (1) rods, plates and connectors, (2) ties and ballast, (3) switching mechanisms and (4) Signal system components. Because the current version is specifically dedicated to measure running surface of turnout including switch points, stock rails, closure rails and frog portions of the turnout [31].

Asplund et al. [32] proposed the idea of using cameras for turnout inspection. Their proposed system is composed of a web-based camera with a minimum resolution of 1600 × 1200 pixels which is protected by a plastic housing. It possess an approximate weight of 3 kg including the internet access module and batteries. The system is mounted above the overhead line to get a fully symmetric bird's eye over the turnout which makes the inspection of both rails and blades as well as geometric calculation possible leading to a reduction in manual inspection frequency. Their proposed system has been tested by the Swedish engineering company, Damill AB. However, the authors have just postulated on how beneficial would be to use cameras for turnout inspection without proving the effectiveness of the proposed system by real in-site measurements.

3.3 Data Collection Tools

The purpose of improving inspection technologies and using new concepts for inspection in turnouts is not solely to simplify the inspection process. Better inspection documentation will provide better understanding of the condition of turnouts over time. This will enable the infrastructure manager to understand degradation processes of turnouts and consequently to predict their maintenance and renewal requirements. This means that a better management of components will be achieved which helps minimizing the maintenance costs and analysing life cycle costs for different maintenance and operation scenarios. Using new technologies for turnout inspection has been rare due to complexity of turnouts themselves. The deterioration of turnouts occurs in two forms of geometrical degradation and tear, wear or plastic deformation. The technologies presented in this paper are developed for geometrical measurements used for predicting geometrical degradation, and also for profile measurements used to plan grinding actions and prevent turnout components from being wore. However, there are technologies which can provide both of them (Table 2).

Table 2 Summary of inspection tools

Name	Function	Maintenance area
SwitchInspect	Inspection documentation- Assigning maintenance indices	Inspection
MiniProf Switch	Providing instant information on metal removal and grinding stone tilt	Grinding
Laser Based Trolley	profile inspection	Grinding-welding
The levelling instrument	Longitudinal level (vertical direction) under unstressed conditions.	Geometrical measurement
Switch Inspection and Measurement (SIM)	Automatic inspection, profile and geometrical measurement	All
Turnout Condition monitoring using machine vision	Detection of tie plates, anchors and spikes	Missing bolts and cotter pins inspection
Health Monitoring of Bolted Joints	Bolted joints detection by piezoelectric transducers and impedance-based	structural health monitoring of bolted joints
Contactless turnout profile measurement	Determining turnout condition in terms of wear and deformation	Grinding
ASIV	Profile and geometry measurement for switch and frog	All
Web-based camera	Profile and geometrical measurement	Automatic inspection

4 Conclusion

Turnouts are among the most critical components of railway track. Their importance lies on their complicated structure and their potential to lead in hazardous accidents. Therefore, considerable budget is being spent on their maintenance and inspection, annually. As a result, it is important to understand the degradation processes of turnouts in order to plan for their maintenance and renewal in advance and to better allocate the budget. In the literature, unavailability of reliable data regarding turnout condition has been an obstacle for understanding their degradation trends. On the one hand, some researchers try to model degradation of turnouts by developing a degradation model based on historical data which, by considering all the influential parameters, is able to predict components degradation in the future. On the other hand, the second approach introduces new inspection technologies providing consistent, objective inspection of turnouts leading to better understanding of their condition and to monitor their degradation. The first approach has had less success as a result of unreliability of required data regarding turnout condition. However, the second approach facilitates inspection and provides a better understanding on turnout degradation trends over time. Although some of such technologies as ASIV or SIM are able to inspect a turnout thoroughly, but some are still under development and need to be improved considering complexities of turnouts.

References

1. Nissen A (2005) Analys av statistik om spårväxlars underhållsbehov. Doctoral dissertation, Samhällsbyggnad, Avdelningen för drift och underhåll, Lulea tekniska universitet
2. Zwanenburg WJ (2009) Modelling degradation processes of switches & crossings for maintenance & renewal planning on the Swiss railway network. Doctoral dissertation, École polytechnique fédérale de Lausanne
3. U.S. Department of Transportation, Federal Railroad Administration (2009) Track Safety Standards Part 213, CFR Title 49
4. Zwanenburg WJ (2007) Degradation processes and wear of railway switches and crossings: the Swiss experience. In: Rail Tech Europe, Seminar 6 (No. LITEP-PRESENTATION-2008-001)
5. Liu X, Saat MR, Barkan CP (2012) Analysis of causes of major train derailment and their effect on accident rates. Transp Res Rec: J Transp Res Board 2289(1):154–163
6. Arasteh khouy I (2013) Cost-effective maintenance of railway track geometry: a shift from safety limits to maintenance limits. Doctoral thesis/Luleå University of Technology
7. Camargo LFM, Edwards JR, Barkan CP (2011, January) Emerging condition monitoring technologies for railway track components and special trackwork. In: 2011 Joint rail conference. American Society of Mechanical Engineers, , pp 151–158
8. Arasteh khouy I, Larsson-Kråik P-O, Nissen A, Lundberg J, Kumar U (2013) Geometrical degradation of railway turnouts—A case study from a Swedish heavy haul railroad. Proc Inst Mech Eng Part F: J Rail Rapid Transit 228(6):611–619. doi:10.1177/0954409713503320
9. Jovanovic S (2004, October). "Railway track quality assessment and related decision making". In: 2004 IEEE international conference on systems, man and cybernetics, vol 6, pp 5038 −5043. IEEE

10. Zarembski AM, Bonaventura CS, Holfeld D (2006, September) Development of maintenance indices for turnouts. In: Proceedings of AREMA conference on railway track and structures
11. MiniProf Switch. https://www.greenwood.dk/miniprofswitch.php. Accessed 15 April 2015
12. Rusu M, Roberts C, Kent S (2012) The use of laser based trolley for railway switch and crossing inspection". In: eMaintenance 2012: the Second international workshop and congress on eMaintenance. Kulturens Hus, Luleå, Sweden, 12th–14th Dec 2012. Luleå University of Technology, Luleå
13. Micro-Epsilon. Laser profile scanner. http://www.micro-epsilon.co.uk/laser-scanner-profilesensor/Laser-scanner-selection/index.html. Accessed 15 April 2015
14. Jönsson J, Khouy IA, Lundberg J, Rantatalo M, Nissen A (2014) Measurement of vertical geometry variations in railway turnouts exposed to different operating conditions. Proc Inst Mech Eng Part F: J Rail Rapid Transit 0954409714546205
15. Eurailscout (2015) Switch inspection & measurement (SIM)—Features and arguments. http://www.eurailscout.com/1-switch-inspection-measurement-sim-features-and-arguments_en.html. Accessed 15 April 2015
16. Molina LF, Resendiz E, Edwards JR, Hart JM, Barkan CP, Ahuja N (2011) Condition monitoring of railway turnouts and other track components using machine vision. In: Transportation Research Board 90th annual meeting (No. 11-1442)
17. Deutschl E, Gasser C, Niel A, Werschonig J (2004, June) Defect detection on rail surfaces by a vision based system. In: 2004 IEEE Intelligent vehicles symposium. IEEE, pp 507–511
18. Mandriota C, Nitti M, Ancona N, Stella E, Distante A (2004) Filter-based feature selection for rail defect detection. Mach Vis Appl 15(4):179–185
19. Popov DV, Titov EV, Mikhailov SS (1999, October) Rail head wear measurements using the CCD photonic system. In: Photonics for transportation. International Society for Optics and Photonics, pp 32–36
20. Aguilar JJ, Lope M, Torres F, Blesa A (2005) Development of a stereo vision system for non-contact railway concrete sleepers measurement based in holographic optical elements. Measurement 38(2):154–165
21. Wamani WT, Villar C (2009) AURORA automated railroad tie condition assessment system: the quest for accuracy. In: 2009 AREMA conference proceedings, American Railway and Maintenance of Way Association (AREMA), Chicago, Illinois
22. Yella S, Dougherty M, Gupta NK (2009) Condition monitoring of wooden railway sleepers. Transp Res Part C: Emerg Technol 17(1):38–55
23. Labarile A, Stella E, Ancona N, Distante A (2004, June) Ballast 3D reconstruction by a matching pursuit based stereo matcher. In: 2004 IEEE Intelligent vehicles symposium. IEEE, pp 653–657
24. Singh M, Singh S, Jaiswal J, Hempshall J (2006, October) Autonomous rail track inspection using vision based system. In: Proceedings of the 2006 IEEE International conference on computational intelligence for homeland security and personal safety. IEEE, pp 56–59
25. Marino F, Distante A, Mazzeo PL, Stella E (2007) A real-time visual inspection system for railway maintenance: automatic hexagonal-headed bolts detection. IEEE Trans Syst Man Cybern Part C: Appl Rev 37(3):418–428
26. Berry A, Nejikovsky B, Gilbert X, Tajaddini A (2008) High speed video inspection of joint bars using advanced image collection and processing techniques. In: Proceedings of the 2008 World congress on railway research, Seoul, Korea
27. Babenko P (2009) Visual inspection of railroad tracks. PhD dissertation. University of Central Florida Orlando, Florida
28. Afshari M, Marquié T, Inman DJ (2009, January) Automated structural health monitoring of bolted joints in railroad switches. In: ASME 2009 rail transportation division fall technical conference. American Society of Mechanical Engineers, pp 1–7
29. Schoone C (2010) Monitoring a turnout of a railway or tramway line. EP 2165915 A2, 23 Sept 2009, 24 Mar 2010

30. Zarembski AM, Palese JW, Euston TL, Scheiring WR (2011) Development and implementation of automated switch inspection vehicle. In: Proceedings of the 2011 Annual conference (AREMA). Minneapolis, MN, 18–21 Sept 2011
31. Zarembski AM (2012) Automated turnout inspection. In: Transportation Research Board 91th Annual meeting, presentation no. p12-6862-1
32. Asplund M, Dan L, Matti R, Arne N, Uday Kumar. (2013) Inspection of railway turnouts using camera. In: World congress on railway research

Part II
Diagnostics, Prognostics and Health Management

Context-Based Maintenance and Repair Shop Suggestion for a Moving Vehicle

Adithya Thaduri, Diego Galar, Uday Kumar and Ajit Kumar Verma

Abstract Maintenance of moving vehicles is quite challenging because they may disrupt the normal flow of transportation due to unexpected breakdowns, slowdowns and stoppages. In order to avoid stoppages and to minimize the downtime, maintenance and condition monitoring systems must be optimized. On one hand the condition monitoring on board should provide automatic failure detection, identification and localization together with a prognostic of the future failures. On the other hand maintenance logistics and product supportability must be also optimized since the onboard system should provide a suggestion of a repair shop that depends on location, cost and availability of spare parts, technicians' skills and queuing time for repairs. However the vehicles are independent assets interacting among them within the traffic system and also interacting with the infrastructure (roads, rails etc.) seriously affected by weather, maintenance of infra, regulations etc. Therefore the proposed solution is to equip the vehicles with a context-aware system that monitors the condition and maintenance schedules of parts and alarm the driver of the parts that are in near to repair cycle. This system will perform risk analysis and will communicate with the cloud propose a decision of selection of repair shop on the location and path of vehicle depending on weather, road and traffic, cost and availability of spare parts at respective repair shops based on risk assessment and prediction. The information contained in the cloud will also communicate the workshop that will book time slot and block the necessary spare parts for the coming vehicle minimizing waiting time. This mechanism will help in reducing unexpected stoppages, vehicle degradation and efficient spare parts management

A. Thaduri (✉) · D. Galar · U. Kumar
Luleå University of Technology, Luleå, Sweden
e-mail: adithya.thaduri@ltu.se

D. Galar
e-mail: diego.galar@ltu.se

U. Kumar
e-mail: uday.kumar@ltu.se

A.K. Verma
Stord/Haugesund University College, Haugesund, Norway
e-mail: akvmanas@gmail.com

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_6

combining in a successful way the workload of the workshops from both natural sources, the time based inspections and repairs together with the reactive maintenance coming from unexpected breakdown.

Keywords Context-aware · Repair shop management · Condition monitoring · Decision support system · Moving vehicle

1 Introduction

Maintenance is one of the driving requirements in operations along with logistics, financing and safety. Especially in railways, it must keep with periodic inspection and maintenance at regular or prior intervals for minimizing the effect of infrastructure failures that may disrupt passenger and cargo services.

If there was a failure in one of the critical components in a moving vehicle which does not surfaced in the corrective maintenance, then it lead to obstruction to ongoing traffic, loss to the asset and human dissatisfaction. The present technologies can able to do several maintenance technologies to detect the ongoing failures but they are not designed with respect to the context and changing environment. For condition monitoring (CM) the components, sensors are installed at several places for data acquisition and then analysed the data for diagnosis and prognosis in a remote location using computerized maintenance management system (CMMS). The combination of CM and CMMS was presently applying in several industries for the effective and efficient prediction of failures in time to reduce the cost, human effort and risk. Computer-based systems are now being used to spontaneously diagnose problems to overcome some of the disadvantages accompanying with relation to experienced personnel [1]. Typically, a computer-based system utilises a linking between the observed symptoms of the failures and the equipment problems using practices such as table look ups [2], symptom-problem matrices, and rules of thumb [3]. These techniques work well for systems with simple mappings between symptoms and problems, but diagnostics seldom have simple correspondences for complex equipment and processes. In addition, not all symptoms are necessarily present if a problem has occurred; making other approaches more cumbersome [4, 5].

Traditionally, the combination of CM and CMMS has been implementing in several systems whether the system is in static or in dynamic motion. There are several challenges in the dynamic motion, for example in railways, the location and the environment is changing rapidly and the systems need to adjust to the environment. There is also need to predict the failure due to continuously dynamic environment and report the possible occurrences of failure to the central computerized system.

For the maintenance activity to be performed for those vehicles in urgent situations there is need for capability to perform replacement or repairs at a maintenance site that is more related to the context of the vehicle. Even though, in general sense, the repair site that is closest to failure site is the best location for the maintenance activity, it may not be always possible because of the other parameters that urge to be considered such as cost, availability, queue, and logistics. In order to overcome it, there is a need of effective diagnostics and prognostics to detect the possible failure of the components in a moving vehicle. Here, moving vehicle may refer to any vehicle that is in transport like automobiles, bus, train, ships or aeroplanes [5].

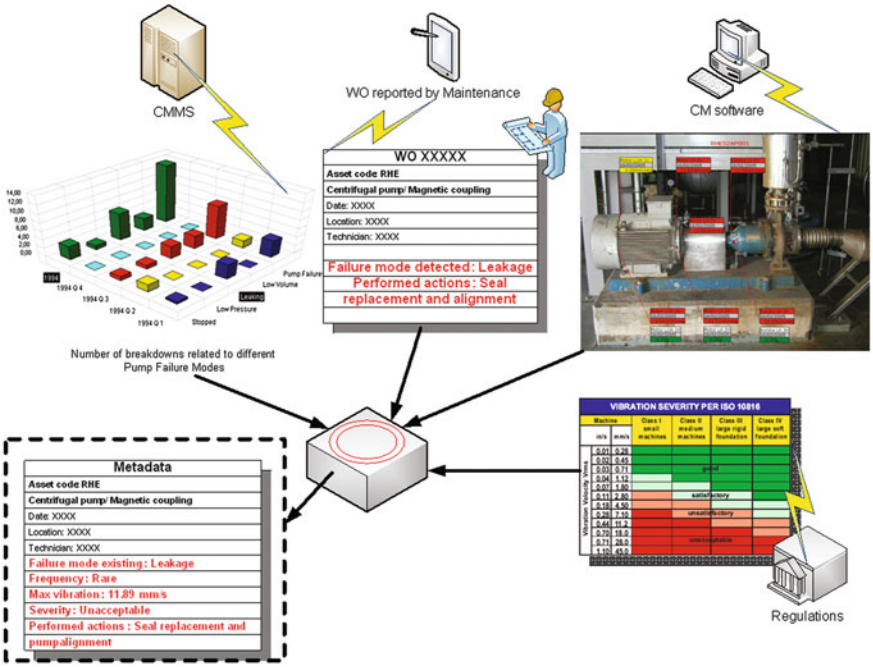
The services for the condition monitoring and computerized maintenance management system can be utilized for effective performance for repair shop management activities. This can be possible by installing on-board diagnostics and prognostics in the moving vehicle for the data acquisition. This onboard system will send data to the centralized cloud system of assisting information [6]. The data in centralized system will do the optimization decision making on the context-driven to provide the best repair shop management and also allocate slot for maintenance activity. The paper is structured as: Sect. 2 gives explanation of maintenance of moving vehicles and how it will perform, Sect. 3 gives information on onboard diagnosis and prognosis that can be installed on moving vehicle, Sect. 4 discusses on the repair shop management and Sect. 5 provides the proposed approach for the onboard maintenance management for repair shop activities.

2 Maintenance of Moving Vehicles

In order to perform maintenance activities, the first step is to collect data from all the components of the vehicle and sent to computerized data management.

2.1 Data Collection

There is a tremendous need to assimilate asset information to get a precise health assessment of the whole system, from various sources such as infrastructure, facilities, factories, vehicles etc., and thereby determine the probability of a shutdown or slowdown, [7]. Moreover, the data acquired are often distributed across independent systems that are challenging to access and not correlated. If the data from these independent systems are combined into a common correlated data source, this rich new set of information could add value to the individual data sources [4]. For example, it is common for most of the facilities to collect work records of where work has been done. Many assets also typically measure their health using condition monitoring (CM) or non-destructive testing (NDT) techniques [8] as “nowcasting” technologies in order to see where work needs to be done. However, these two datasets can remain in separate and individual systems. By combining the data into a location correlated dataset, i.e. metadata (Fig. 1), the quality and/or the effectiveness of the work being performed



WO XXXXX

Asset code RHE

Centrifugal pump/ Magnetic coupling

Date: XXXX

Location: XXXX

Technician: XXXX

Failure mode detected: Leakage

Performed actions: Seal replacement and alignment

Metadata

Asset code RHE

Centrifugal pump/ Magnetic coupling

Date: XXXX

Location: XXXX

Technician: XXXX

Failure mode existing: Leakage

Frequency: Rare

Max vibration: 11.89 mm/s

Severity: Unacceptable

Performed actions : Seal replacement and pumpalignment

VIBRATION SEVERITY PER ISO 10816

Machine		Class I small machines	Class II medium machines	Class III large rigid foundation	Class IV large soft foundation
in/s	mm/s				
0.01	0.28	good	good	good	good
0.02	0.45	good	good	good	good
0.03	0.71	good	good	good	good
0.04	1.12	satisfactory	good	good	good
0.07	1.80	satisfactory	satisfactory	good	good
0.11	2.80	unsatisfactory	satisfactory	satisfactory	good
0.18	4.50	unsatisfactory	unsatisfactory	satisfactory	satisfactory
0.28	7.10	unacceptable	unsatisfactory	unsatisfactory	satisfactory
0.44	11.2	unacceptable	unacceptable	unsatisfactory	unsatisfactory
0.70	18.0	unacceptable	unacceptable	unacceptable	unsatisfactory
0.71	28.0	unacceptable	unacceptable	unacceptable	unacceptable
1.10	45.0	unacceptable	unacceptable	unacceptable	unacceptable

Vibration Velocity Vrms

Fig. 1 Metadata for maintenance knowledge extraction

can be analysed by comparing the “asset health” before and after the work is completed [4, 5, 9].

Figure 2 shows the systems currently used by maintainers in factories or facilities. Computerized maintenance management system (CMMS) and CM are the most popular repositories of information in maintenance, where most of the deployed technology is installed and unfortunately isolated information islands are usually created [10]. While using a good version of either technology can assist in reaching the defined maintenance goals, combining the two (CMMS and CM) into one seamless system can have exponentially more positive effects on maintenance and asset performance than either system alone might achieve. The combination of the strengths of a top-notch CMMS (preventive maintenance (PM) scheduling, automatic work order generation, maintenance inventory control, and data integrity) with the capabilities of a leading-edge CM system (multiple-method condition monitoring, trend tracking, and expert system diagnoses) in such a way that work orders are generated automatically based on information by CM diagnostic and prognostic capabilities improving dramatically the asset performance, [5, 11–13].

Just a few years ago, linking CMMS and CM technology was mostly a vision easily dismissed as infeasible or at best too expensive and difficult to warrant much investigation. Now, due to the advancement of computing technologies, there is a possibility that the combination of CMMS and CM have been to implement to

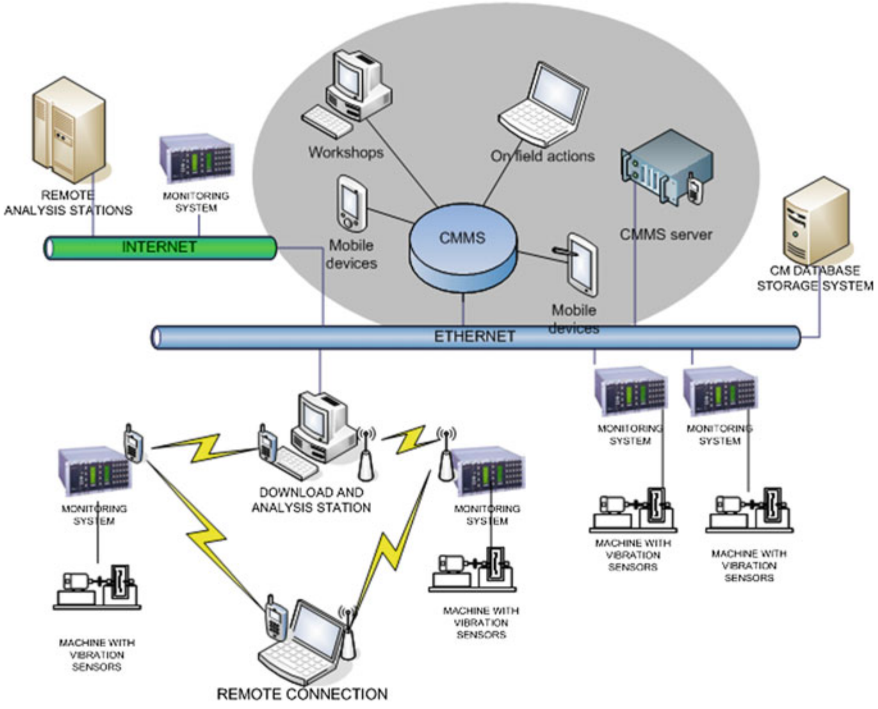


Fig. 2 ICT architecture for the integration of CMMS and CM systems in maintenance and asset management

achieve such a link relatively easily and inexpensively. A top-shelf CMMS can perform a wide variety of functions to improve maintenance performance, [10]. It is the central organizational tool for World-Class Maintenance (WCM). Among many other critical features, a CMMS is primarily designed to facilitate a shift in emphasis from reactive to preventive maintenance. It achieves this shift by allowing maintenance professional to set up automatic PM work order generation. A CMMS can also provide historical information which is then used to adjust PM system setup over time to minimize unnecessary or redundant maintenance actions or repairs, while still avoiding run-to-failure repairs. PMs for a given piece of equipment can be set up on a calendar schedule or a usage schedule that utilizes meter readings. A fully-featured CMMS also includes inventory tracking, logistics, workforce management, purchasing, in a package that stresses database integrity to safeguard vital information [14]. The final result is optimized equipment up-time, lower maintenance costs, and better overall plant efficiency [5].

On the other hand, a CM system should accurately monitor real-time equipment performance, and alert the maintenance professional to any changes in performance trends. There are a variety of measurements that a CM package might be able to track including vibration, oil condition, temperature, operating and static motor characteristics, pump flow, and pressure output [15]. These measurements are squeezed out of equipment by monitoring tools like Ferro graphic wear particle analysis, proximity probes, triaxle vibration sensors, accelerometers, lasers, and multichannel spectrum analysers [14]. The very best CM systems are expert systems that can analyse measurements like vibration and diagnose machine faults [5, 11, 13].

2.2 Context Driven Maintenance Decisions

A context-aware system actively and autonomously adapts and provides the most appropriate services or information to users, taking advantage of people's contextual information while requiring little interaction. The concept of context-aware computing was quoted by [16, 17] as "the ability of a mobile user's applications to discover and react to changes in the environment they are situation" [18].

Context-aware systems are usually complicated and are responsible for many jobs, such as representation, modelling, management, reasoning, and analysis of context information. They require the collaboration of many different components in the systems. There are various types of different context-aware systems, making it hard to generalise a context-aware system process; however, a context-aware system usually follows four steps as shown in Fig. 3 [9].

The first step is acquiring context information from sensors. Sensors convert real world context information into computable context data. By using physical and virtual sensors, the system can capture various types of context-aware information. The system then stores the data into its repository. When storing context data, the kind of data model used to represent the context information is very important; context models are diverse, and each has its own unique characteristics. To easily use the stored context data, the system controls the abstraction level of the data by interpreting or aggregating them. Finally, the system uses the abstracted context data for context-aware applications. One such representation of context aware system was developed by [19] for intelligent broker systems. They developed context broker architecture (CoBrA) for context acquisition using different sensors with devices among users, machines and agents to provide adequate support for context modeling by analyzing from data repositories. The features are then extracted for the application of brokerage activities.

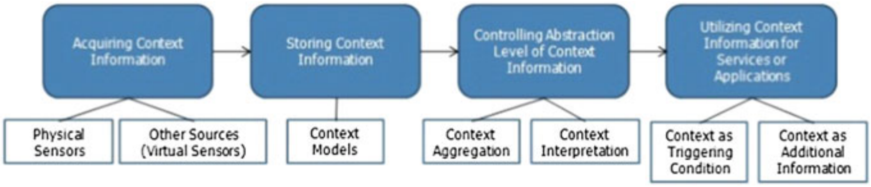

Acquiring Context Information
Storing Context Information
Controlling Abstraction Level of Context Information
Utilizing Context Information for Services or Applications
Physical Sensors
Other Sources (Virtual Sensors)
Context Models
Context Aggregation
Context Interpretation
Context as Triggering Condition
Context as Additional Information

Fig. 3 General process in context-aware systems

2.3 Context Driven Condition Monitoring

In the past, the different functional areas, e.g., the process monitoring, the equipment monitoring and the performance monitoring, were performed independently and each tried to "optimize" their associated functional area without regard to the effect that given actions might have on the other functional areas [20]. As a result, a low priority equipment problem may have been causing a large problem in achieving a desired or critical process control performance, but was not being corrected because it was not considered very important in the context of equipment maintenance. With the asset cloud providing data to the end users, however, persons can have access to a view of the plant based on two or more of equipment monitoring data, process performance data, and process control monitoring data. Similarly, diagnostics performed for the plant may take into account data associated with process operation and the equipment operation and provide a better overall diagnostic analysis.

Due to advent of advancement in computing and data acquisition capabilities, the competences of condition monitoring is driven across several fields [21]. As in respect to the context-aware systems, the condition monitoring techniques is applying at several cross-board areas with abundant enhancements to the machines and users with effectiveness and efficiency. Due to dynamic and adaptive environments, the context-aware condition monitoring helps in reducing the risks, safety by effective remaining useful life prediction for condition based maintenance [22].

2.4 Diagnosis with Anomaly Detection

The anomaly detection task is to recognize the presence of an unusual (and potentially hazardous) state within the behaviours or activities of a system, with respect to some model of 'normal' behaviour which may be either hard-coded or learned from observation [23]. We focus, here, on learning models of normalcy at the user behavioural level, as observed. An anomaly detection agent faces many learning problems including learning from streams of temporal data, learning from instances of a single class, and adaptation to a dynamically changing concept [24]. In addition, the domain is complicated by considerations of the trusted insider problem (recognizing the difference between innocuous and malicious behaviour changes on the part of a trusted user) and the hostile training problem (avoiding learning the behavioural patterns of a hostile user who is attempting to deceive the agent).

Anomaly detection discusses to the problem of discovery of patterns in data that do not follow to predictable behaviour. These non-conforming patterns are often denoted as anomalies, aberrations, contaminants, discordant, exceptions, observations, outliers, peculiarities or surprises in different application domains. Noise removal is determined by the necessity to eliminate the undesirable objects formerly any data analysis is accomplished on the data [23, 25].

Contextual anomalies have been most frequently reconnoitred in time-series data and spatial data. The selection of relating to a contextual anomaly detection technique is definite by the significance of the contextual anomalies in the application domain. Another main issue is the accessibility of contextual attributes. In some cases defining a context is direct, and hereafter relating a contextual anomaly detection technique makes sense [26]. In other cases, defining a context is difficult, making it hard to apply such techniques [23].

2.5 Context-Driven E-Maintenance

Once connectivity is sorted out then sense making becomes the real challenge for data sets. It is therefore time for migrating concepts from e(lectronic) Maintenance to i(ntelligent) Maintenance, [27]. Maintainers must deal many different sources of information. In this paper, we use a system framework supporting the integration of various data sources which could have different formats and natures. To handle those differences, the system framework should provide facilities for data wrapping and mediation between different data formats, along with interfaces for external data wrappers and mediators. The system should also be able to add new sources and mediation procedures and handle the necessary data validation and consistency checking. From the operation point of view, different data spaces must be managed at different levels of the system. At the data management data space, the following agents and databases must be managed and merged for Database, containing the database baseline [11];

- Synthetic database, containing derived calculations from the database or from external sources not included in the database;
- Information on managing the databases;
- Information on managing wrappers and mediators;
- Archived data.

2.6 Prognosis for Health Assessment

Beside safety hazards, there are two basic risks associated with assets: shutdowns and slowdowns. These risks materialise in economic loss, [28]. The only way to save money is to perform a proper prognosis, not just a diagnosis [4, 11]. The monitoring equipment depends on many sets of instruments or sensors which are suitably distributed to obtain information about system state. For this reason the monitoring activity represent a key point of the whole system under examination. An erroneous feedback due to instrument failure may cause damages whose extent depends on the control system sensitivity to the incorrect measurements. In fact, if you are not measuring something accurately and consistently, you do not know if your inferences are valid. These aspects are of fundamental importance in all those fields where the reliability of data measured by sensors or, more in general, by instruments has to be assured before using them for implementing subsequent actions. Moreover if an instrument failure occurs this may leads also to false or missing actions so giving safety problems. There are three basic ways to model how faults develop: symbolic models, data-driven models, and physics of failure models based on physical principles, laboratory tests and measurements and mathematical formulations [11].

3 Onboard and Centralized Diagnosis and Prognosis

The communication system consists of both onboard devices that are installed in the moving vehicle and a centralized data system to perform critical data analysis.

3.1 Autonomous Onboard System

Onboard devices and applications can be used for model based knowledge representation for the existing systems for autonomy. In the event of failure operation, it is difficult to find the component failed due to unknown reasons that were not surfaced in the inspection and condition monitoring. In such cases, there is a necessity of an autonomous onboard system that is not only monitors the several sub systems in the vehicle and also is capable of sending the information to the cloud for diagnosis and prognosis [29]. This system will acquire data from the all the components using sensors. Integrated onboard and centralized reasoning systems capable of blending results from multiple sensors and driver to be informed the health of the vehicle needs to be applied. This engine and the test procedures have to be solid enough so that they can be embedded in the electronic control unit (ECU) and/or a diagnostic maintenance computer. Due to on site, the response time can also be faster than the data analysis. Usually, the human (driver) are responsible for the request or report the problem to the maintenance system. In the case of this autonomous onboard maintenance system, there is no need of human to involve. The transceiver present in the board will send signals to centralized system automatically if any one of the components will reach near to its maintenance activity. Several algorithms that interface with onboard usage monitoring systems and parts management databases are used to predict the useful life remaining of system components for maintenance activities [30].

The transceiver also capable of receiving request from the centralized system to inform the driver about possible decision making that is performed in the cloud. The extension markup language (XML) transformation of onboard data and the protection of context based on location with meta-data techniques, automated test meta-language (ATML) based tests and results and diagnostic result encapsulation data can be pooled as inputs for the presentation of data mining techniques. The applicability of the data collected early in the diagnostics and maintenance process is performed effectively through in this concept [31]. The functions of the onboard system are:

- To perform diagnosis and prognosis of several components in the moving vehicle.
- To store data from several maintenance records to keep track of its repair/replacement cycles.
- To send the possible reaching of component's maintenance activity or failure propagation that is observed by condition monitoring.
- To receive the requests from the centralized system and inform the user.
- To inform the user about possible condition of the moving vehicle onboard without centralized system.

3.2 Centralized System

The data from the onboard devices from several moving vehicles is fed to a centralized system that handles huge amount of data. The data is collected by two ways. The large amount of condition monitoring data and updated maintenance activities can be uploaded to centralized system either by moving vehicle reaching the centralized data system or user can download data from the device separately and upload through internet connection. The small amount of data like requests, alerts, notifications or any other data can be send through General packet radio service (GPRS) mobile data using 2G or 3G services that is installed on the device. The data analysis, model based methods, techniques for fault diagnosis, prognosis techniques, interpretation of the possible scenarios, update of sources can be performed at centralized system instead of onboard [32] since these techniques need huge amount of processing information and data. There are several advancements in this area of diagnosis and prognosis with combination of condition monitoring and condition based maintenance with e-maintenance by usage of cloud and other services. This system then further provide decision making of the respective repair shop suggestions based on context of the moving vehicle and logistics of repair shops.

4 Repair Shop Management

For many industries, especially for expensive and risk based assets, repairing a failed asset is significantly more economical than replacing it and in some conditions, companies often cannot even afford inventory. Hence, in such conditions, a high-quality scheduling, logistics along with maintenance activities is incorporated to improve the performance of the overall system. The placement of the repair shop was planned by considering factors like location, availability of the components, cost effective, policies [33], logistics, number of consumers and affordability. Due to the variable need of demand and supply, the items are in general maintained the inventory or transport the items that are rarely went to failure. There are basically three major factors considered for the performance of the environment as [34]:

1. The initial spares inventory levels for final assemblies, subassemblies, and components.
2. The capacity to repair parts and to perform inspection, assembly, and testing of subassemblies and final assemblies.
3. The priority scheduling system used in the repair shop.

There were several researches going on these factors with advancement in computing. Accordingly, the progress of proper overtime policies for a repair environment need of attention of several issues related to any job shop environment. Five areas will be discussed here [33],:

1. The fundamental trade-off involved
2. When overtime should be used
3. How much overtime to use
4. What level in the product structure to work overtime
5. Job and labour scheduling policies

Even in the repair shop, there are several disruptions that can happen as listed in Table 2 [35]. These disruptions are complex and require specialised repair. When the centralized system “talked” to the repair shop, it must provide the present condition of the repair shop, the status of the inventory, the scheduling queue and logistics.

5 Proposed Cloud-Based Repair Management

This paper proposes a conceptual network of onboard diagnosis and prognosis with condition monitoring, centralized CMMS cloud infrastructure and repair shop management that provide decision based on context-driven as shown in Fig. 4. Even though similar studies are implemented in aircrafts, this approach has been transferred to suit the moving vehicles. The novel part is the condition monitoring of the onboard diagnosis and how different it works based scheduling, communication, data transfer and risk assessment. The process of this methodology is explained in following steps.

(1) The onboard device will detect the possible maintenance activity based on diagnosis and prognosis or maintenance cycle of the components.
(2) The request for the maintenance activity is sent to the nearest centralized system..

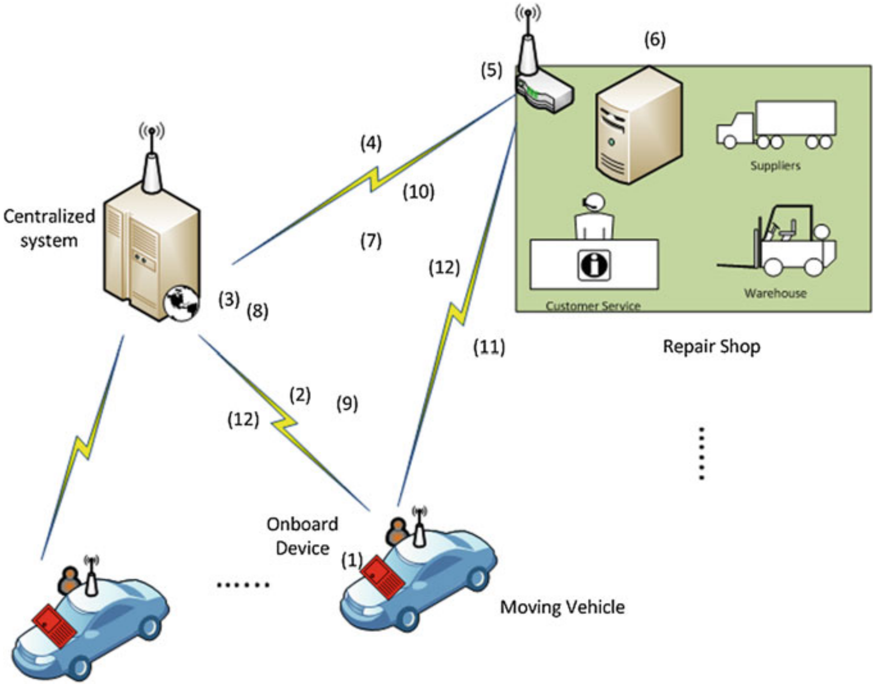


Fig. 4 Proposed on-board-centralized system for repair shop suggestion

(3) The centralized system will perform several analysis and artificial intelligence tools to approve and accept the request generated by the onboard system. The system also looks for nearest repair shop based on location.

(4) The requests consists of part, time remained, cost and other logistics details is sent to several repair shops and will wait for the response.

(5) The requests sent by the centralized system will be received and analyse the request.

(6) The request is then look for several factors like cost, availability, queue, maintenance personnel, time taken, location along with disruption shown in Table 1.

(7) The above information is again send back to centralized station for approval process.

(8) The centralized system then analyses the requests from several repair shops and provides decision support system based on logistics and select the optimized solution of repair shop.

(9) The information of selected repair shop is sent to the requested moving vehicle.

(10) The information of selected repair shop is also sent to the respective repair shop for confirmation.

Table 1 Disruptions on the repair shop

Sl. no	Disruption
1	Machine breakdown
2	Maintenance of machine
3	Absenteeism
4	Tool breakdown
5	Process time variation
6	Delay in transport using material handling system
7	Variation in performance of machine
8	Tool wear
9	Variation of set-up times
10	Arrival of a new job order
11	Rework
12	Rejection
13	Unavailability of raw material
14	Urgent job
15	Change of priority
16	Cancellation of order
17	Outsourcing

(11) The repair shop then communicates with moving vehicle in critical condition and will do the job.

(12) Once the job is completed, the maintenance activity is stored in the centralized database for future revisions.

There is prerequisite of future work that involves the following functions

- the communication protocols and software implementation among onboard device, centralized system and repair shop
- the factors and parameters required for data storage in all the three systems
- the selection and implementation of several algorithms required to provide decisions for diagnosis, prognosis, condition monitoring and suggestion of repair shop.

6 Conclusion

Even though due to the advancement of several technologies, there is need of applying these concepts in maintenance activities to reduce risk, cost, manage logistics and burden in practical field. One of the fields that need more concentration is the effective utilization of these technologies in repair shop management. This conceptual paper proposes a novel approach for implementation of such case. This paper studies the several maintenance activities available in the literature, the on-board devices that can be incorporated, the functions of centralized system and application of all these technologies to provide suggestion for the best repair shop with consideration of logistics and context-driven mechanisms. There is need of several advancements in this context to provide our knowledge to improve the performance of machine and human in maintenance area.

References

1. Price C, Price CJ (1999) Computer-based diagnostic systems. Springer, Heidelberg, pp 65–69
2. Gandolfo F, Mussa-Ivaldi FA, Bizzi E (1996) Motor learning by field approximation. Proc Natl Acad Sci 93(9):3843–3846
3. Fischhoff B, Slovic P, Lichtenstein S (1978) Fault trees: Sensitivity of estimated failure probabilities to problem representation. J Exp Psychol Hum Percept Perform 4(2):330
4. Galar D, Kumar U, Villarejo R, Johansson CA (2013) Hybrid prognosis for railway health assessment: an information fusion approach for PHM deployment. Chem Eng 33
5. Galar D, Thaduri A, Catelani M, Ciani L (2015) Context awareness for maintenance decision making: a diagnosis and prognosis approach. Measurement
6. Rizzoni G, Onori S, Rubagotti M (2009, June) Diagnosis and prognosis of automotive systems: motivations, history and some results. In: Proceedings of the 7th IFAC Symposium on fault detection, supervision and safety of technical processes (SAFEPROCESS'09)
7. Galar D, GuSTAFSON A, Tormos B, Berges L (2012) Maintenance decision making based on different types of data fusion Podejmowanie Decyzji Eksploatacyjnych W Oparciu O Fuzję Różnego Typu Danych. Eksploatacja i Niezawodnosc, Maint Reliab 14(2):135–144
8. Paipetis AS, Matikas TE, Aggelis DG, Van Hemelrijck D (eds) (2012) Emerging technologies in non-destructive testing V. CRC Press, Boca Raton
9. Galar D (2014) Context-driven maintenance: an eMaintenance approach. Manag Syst Prod Eng. http://wydawnictwo.panova.pl/pliki/15_2014/2014_03_05_GALAR.pdf
10. Labib AW (2004) A decision analysis model for maintenance policy selection using a CMMS. J Qual Maint Eng 10(3):191–202
11. Galar D, Palo M, Van Horenbeek A, Pintelon L (2012) Integration of disparate data sources to perform maintenance prognosis and optimal decision making. Insight-non-destructive testing and condition monitoring 54(8):440–445
12. Muller A, Marquez AC, Iung B (2008) On the concept of e-maintenance: review and current research. Reliab Eng Syst Saf 93(8):1165–1187
13. Van Horenbeek A, Pintelon L, Galar D Integration of disparate data sources to perform maintenance prognosis and optimal decision making. http://pure.ltu.se/portal/files/40110675/317_Horenbeek.pdf
14. Bjorling SE, Baglee D, Galar D, Singh S (2013) Maintenance knowledge management with fusion of CMMS and CM
15. Nandi S, Toliyat HA, Li X (2005) Condition monitoring and fault diagnosis of electrical motors-a review. IEEE Trans Energy Convers 20(4):719–729
16. Schilit BN, Theimer MM (1994) Disseminating active map information to mobile hosts. IEEE Netw 8(5):22–32
17. Schilit B, Adams N, Want R (1994, December) Context-aware computing applications. In: WMCSA 1994. First workshop on mobile computing systems and applications, 1994. IEEE, pp 85–90
18. Thaduri A, Kumar U, Verma AK Computational intelligence framework for context-aware decision making. Int J Syst Assur Eng Manag 1–12
19. Chen H, Finin T, Joshi A (2003) An intelligent broker architecture for context-aware systems. PhD proposal in computer science, University of Maryland, Baltimore, USA

20. Nixon M, Keyes M, Schleiss T, Gudaz J, Belvins T (2001) U.S. Patent Application 09/953,811
21. Remboski D, Brooks K, Canavan P, Douros K, Gardner J, Gardner R, Hurwitz J, Leivian R, Nagel J, Wheatley D, Wood C (2001) U.S. Patent Application 09/976,974
22. Bottazzi D, Corradi A, Montanari R (2006) Context-aware middleware solutions for anytime and anywhere emergency assistance to elderly people. IEEE Commun Mag 44(4):82–90
23. Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. ACM Comput Surv (CSUR) 41(3):15
24. Lane TD (2000) Machine learning techniques for the computer security domain of anomaly detection
25. Ye N, Chen Q (2001) An anomaly detection technique based on a chi-square statistic for detecting intrusions into information systems. Qual Reliab Eng Int 17(2):105–112
26. George J, Crassidis J, Singh T, Fosbury AM (2011) Anomaly detection using context-aided target tracking. J Adv Inf Fusion 6(1):39–56
27. Galar D, Wandt K, Karim R, Berges L (2012) The evolution from e (lectronic) maintenance to i (ntelligent) maintenance. Insight-Non-Destr Test Cond Monit 54(8):446–455
28. Wilmering TJ, Ramesh AV (2005, March) Assessing the impact of health management approaches on system total cost of ownership. In: 2005 IEEE Aerospace conference. IEEE, pp 3910–3920
29. de Novaes Kucinskis F, Ferreira MGV (2008) An Internal State Inference Service for onboard diagnosis, prognosis and contingency planning applications
30. Luo J, Pattipati KR, Qiao L, Chigusa S (2007) An integrated diagnostic development process for automotive engine control systems. IEEE Trans Syst, Man Cybern Part C: Appl Rev 37 (6):1163–1173
31. Byington CS, Kalgren PW, Dunkin BK, Donovan BP (2004, March) Advanced diagnostic/prognostic reasoning and evidence transformation techniques for improved avionics maintenance. In: 2004 IEEE Aerospace conference, 2004. Proceedings, vol 5. IEEE
32. Sankavaram C, Pattipati B, Kodali A, Pattipati K, Azam M, Kumar S, Pecht M (2009, August) Model-based and data-driven prognosis of automotive and electronic systems. In: IEEE International conference on automation science and engineering, 2009. CASE 2009. IEEE, pp 96–101
33. Scudder GD (1985) An evaluation of overtime policies for a repair shop. J Oper Manag 6 (1):87–98
34. Scudder GD, Hausman WH (1982) Spares stocking policies for repairable items with dependent repair times. Naval Res Logist Q 29(2):303–322
35. Subramaniam V, Raheja AS (2003) mAOR: A heuristic-based reactive repair mechanism for job shop schedules. Int J Adv Manuf Technol 22(9–10):669–680

Optimal Sensor Placement for Efficient Fault Diagnosis in Condition Monitoring Process; A Case Study on Steam Turbine Monitoring

Farzin Salehpour Oskouei and Mohammad Pourgol-Mohammad

Abstract Failure root cause analysis requires an optimum sensor network in the process of a complex system monitoring. Selection of the location, type and number of sensors are important metrics of sensor network optimization. Main aspects of this optimization can be categorized to failure detection, failures diagnosis from each other, the collected data from sensors and sensor reliability. In the process of sensor networks optimization, logical relationships are determined between components and sub-systems through different methods such as FMEA, FTA and RBD. In this paper, an augmented FMEA and FTA method is developed to extract for predicting failure causes in a condition monitoring process. The potential location of sensors is first determined through Sensor Placement Index (SPI). SPI depends on the Importance of failure modes and the cost of their monitoring processes. Due to the potential places of sensors, different scenarios are derived for sensor placement. Considering prior information about component state (operational or failed), system is simulated through Bays Monte Carlo method. By estimation of sensor detection probability, posterior probability of failure modes is calculated. Then the variance of proposed probabilities is added together and the result represents the uncertainty index. For determining the sensor reliability index, sensors are considered as system components. In this case, functional model of each scenario is developed and the scenario with less Top Event probability is selected as the optimal one. The main purpose of this paper is to show the difference between prioritization of scenarios based on two proposed criterion. It represents that both the uncertainty and reliability of sensors must be considered in the optimization process. But in some specific cases such high-reliable systems, the effect of sensor

F.S. Oskouei (✉) · M. Pourgol-Mohammad
Department of Mechanical Engineering, Sahand University of Technology,
Tabriz, East Azerbaijan, Iran
e-mail: Farzin.salehpour@gmail.com

M. Pourgol-Mohammad
e-mail: pourgolmohammad@sut.ac.ir

F.S. Oskouei
Department of Mechanical Engineering, Shabestar Branch-Islamic
Azad University, Tabriz, East Azerbaijan, Iran

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_7

reliability index can be negligible. As a case study, optimization of sensor placement has been demonstrated on steam turbine and results are discussed.

Keywords Optimal sensor placement · Condition monitoring · Steam turbine · Sensor reliability · Uncertainty

1 Introduction

Increasing operation and maintenance cost has caused more technical interest in mechanical systems on as-needed maintenance methods such as condition-based approaches instead of inefficient scheduled alternatives [1]. In proposed methods, future failures of the system are predicted based on current state of its components. It is clear that failure root cause analysis requires an optimum sensor network design in the process of a complex system monitoring. The location, type and number of sensors are important metrics of sensor network optimization [2–4]. Main aspects of this optimization can be categorized to failure detection, failures diagnosis from each other and the collected data from sensors.

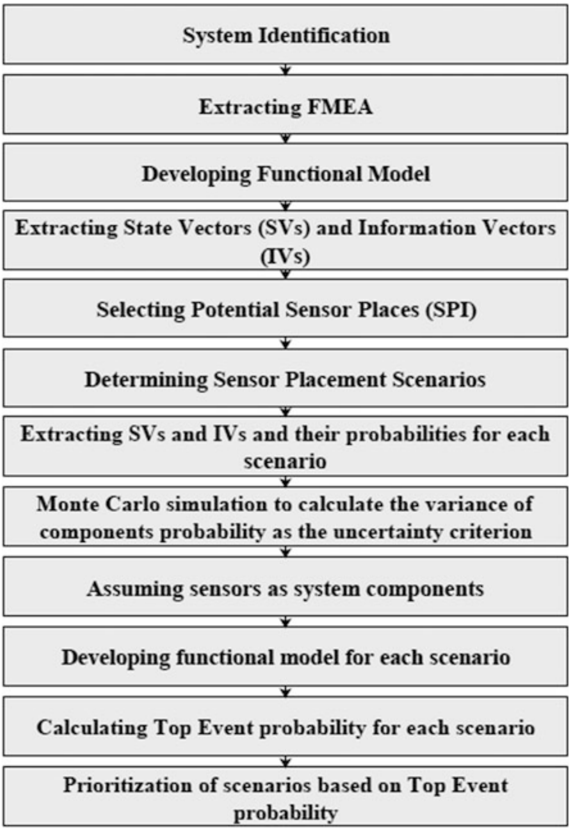
There are different researches about fault diagnosis and condition-based maintenance of mechanical systems [1–7]. Few of proposed studies consider the impact of sensor network on the fault diagnosis process. Obviously, data collection on the state of components will have a significant influence on the reliability of predictions. The performance of a sensor network can be identified by four indicators consisting of fault detection, fault diagnosis, reliability of sensors and data obtained from sensors. In the decision-making process, more reliable information is obtained by reducing the uncertainty of primary hypothesis.

The techniques which are used for sensor placement optimization are mostly focused on finding the optimal physical location of sensors, given some geometrical constraints [8–16]. Proposed methods are based on the Fisher information matrix. The Fisher information in statistical mathematics is a method for measuring visible random variable information about an unknown parameter. In fact, this matrix represents the variance of outcomes or expected values of observed data. In this method, the whole structure is meshed and information matrix is developed for different nodes. Then using an optimization method, the node with a maximum determinant of Fisher information matrix is selected for sensor placement. Another category of optimal method is based on optimizing a cost function considering the constraints of fault detection, fault diagnosis and reliability of sensors [17–22]. Some methods are focused about the probabilistic aspect of sensor placement process [23–25]. Bayesian theory is applied in such methods to extract the posterior information based on historical data. Then deviation of posterior data is calculated through utility function. Considering supposed deviation, prioritization is determined for potential sensor placement scenarios [25].

According to the literature, the main technical interest is about information uncertainty based on prior data, collected by sensor network [24, 25]. However an attention was not paid on reliability of sensors and its impact on optimization process. Also selection of potential places for sensors was not discussed. In this paper, the main motivation is to define an index for prioritization of potential places of sensors. Also considering sensors as components of the system, effect of sensor reliability is studied on optimal sensor placement.

The organization of the paper is managed as follow; in Sect. 2, a functional model is developed for the system. Also the state of each component is extracted as the State Vector (SV). Using SVs, collected information has been arranged from sensors in the form of Information Vector (IV). Based on the uncertainty of proposed information, the optimal placement has been selected for sensors. In Sect. 3, the effect of sensor reliability is studied. Sensors are considered as system components and the optimal scenario is selected based on proposed criterion. Finally in Sect. 4, difference of scenario prioritization has been discussed in both categories. Methodology structure is illustrated in Fig. 1.

Fig. 1 Methodology Structure


System Identification
Extracting FMEA
Developing Functional Model
Extracting State Vectors (SVs) and Information Vectors (IVs)
Selecting Potential Sensor Places (SPI)
Determining Sensor Placement Scenarios
Extracting SVs and IVs and their probabilities for each scenario
Monte Carlo simulation to calculate the variance of components probability as the uncertainty criterion
Assuming sensors as system components
Developing functional model for each scenario
Calculating Top Event probability for each scenario
Prioritization of scenarios based on Top Event probability

2 Developing a Model for System State Diagnosis

For developing an optimization algorithm for the sensor network arrangement, it is necessary to study the complete system and its components from the intended scope and objectives. Also it needs to consider the failure data of each component. Based on these requirements, seven steps are developed for optimal sensor placement algorithm which will be discussed in following sections.

Step 1: First step in sensor placement optimization contains extracting components of the system and their failure modes. This step is performed by applying Failure Modes and Effect Analysis (FMEA) method. Using proposed method, in addition to diagnose failure modes and their effects, importance of each mode is calculated through Risk Priority Number (RPN).

Step 2: In this step, functional model of the system is developed. Different methods such as Fault Tree Analysis (FTA) or Reliability Block Diagram (RBD) can be used to model the logical relation between different components and failure modes. Since the sensor placement problem is directly related to the system operation, the effect of developing an appropriate functional model on optimization process become clearer.

Step 3: Potential locations of sensors are defined in this step. A criterion is specified in this research to reflect both effect of each components failure on the system failure and monitoring costs. Sensor Placement Index (SPI) is defined for each system component by Eq. (1) as:

$$SPI = \frac{Reliability\,Importance}{Monitoring\,Cost} \tag{1}$$

Using Birnbaum Method [26], the Reliability Importance (RI) of each component is determined by Eq. (2) as:

$$I_i^B = \frac{\partial R_S[R(t)]}{\partial R_i(t)} \tag{2}$$

$R_S[R(t)]$, as the reliability equation of whole system, is dependent of each component reliability ($R_i(t)$). Ranking system components (based on SPI), important components are extracted. These components are considered as candidates of potential places for sensors. According to the sensor quantity which is specified based on cost and placement constraints, Number of potential places will be considered. The optimization problem will be meaningful if the number of potential places be more than sensor quantities.

Step 4: Failure modes in the lowest level of system are considered as inputs of the model. So all possible combinations of inputs' states are determined as system state vectors (SV). The state vector represents the occurrence or

Table 1 System state vectors

State vector	SV1	SV2	SV3	SV4	SV5	SV6	SV7	SV8
Failure mode 1	1	1	1	0	1	0	0	0
Failure mode 2	1	1	0	1	0	1	0	0
Failure mode 3	1	0	1	1	0	0	1	0

non-occurrence of system failure modes. One and zero are used to indicate occurrence and non-occurrence of each failure mode respectively. For a system with n failure modes, there are 2^n state vectors. As an example, state vectors of a system with 3 failure modes are shown in Table 1:

In Table 1, SV1 represents occurrence of all failure modes whereas SV8 represents a state in which none of failure modes were occurred. Given the primary occurrence probability of failure modes, occurrence probability of each state vector is calculated through Monte Carlo simulation [24]. It is clear that the summation of all state vectors' occurrence probability must be equal to 1.

Step 5: Considering both available sensor quantities and potential places for them, placement scenarios are developed. If there are p quantity of sensors and m potential places (p < m), then the number of scenarios is calculated through Eq. (3):

$$C(p,m) = \frac{m!}{p!(m-p)!} \tag{3}$$

Each scenario contains different information about system state. To specifying these differences, Information Vectors (IVs) are determined [24, 25] indicating the state of sensors. Sensor state is determined in the binary form where 1 means existence of an alarm and zero means no alarm. Considering p sensors for each scenario, there are 2^{p} IVs. As an example, IVs for a scenario with 3 sensors are shown in Table 2:

For estimating the occurrence probability of each information vector, state vectors are extracted based on the occurrence of related IV. Then probabilities of supposed state vectors are added together and the result indicates the probability of proposed IV.

Step 6: In this step, considering the occurrence probability of information vectors as prior information, posterior state vectors are reproduced through Monte Carlo simulation. According to the posterior SVs, occurrence probability

Table 2 System information vectors

Information vector	IV1	IV2	IV3	IV4	IV5	IV6	IV7	IV8
Sensor 1	1	1	1	0	1	0	0	0
Sensor 2	1	1	0	1	0	1	0	0
Sensor 3	1	0	1	1	0	0	1	0

of each failure mode will be calculated. More description about this step will be given in the section of case study.

Step 7: Information uncertainty is discussed for previous step results in this step. Standard deviation of each failure mode's occurrence probability is calculated for the number of iterations in Monte Carlo process. The inverse of variance is calculated for all failure modes and the summation of them is considered as an index for comprising different scenarios from the uncertainty point of view [24]. Applying proposed index, the ability of each scenario can be measured in detecting system failures.

By completing these steps, sensor placement scenarios will be prioritized based on information uncertainty. However, the reliability of sensors is not considered in this process. False alarm or missed-alarm can cause misunderstanding of system state. So it is necessary to study the effect of sensor failure on optimization process.

3 Effect of Sensor Reliability on Optimal Sensor Placement

As it is discussed, sensor is a crucial component in condition monitoring process. The validity of sensor information and ensuring of its accuracy is a concern in such processes. So the effect of sensor failure is considered on optimization process in this research.

According to the proposed algorithm for sensor placement, one alternative approach is to consider each sensor as a component of the system in order to study the effect of sensor reliability. In condition monitoring process, while the failure of component is detected by a sensor, it doesn't count as a failure. In this case, sensor and the component, monitored by that sensor, will be added to the system model in the parallel form. Thus when both sensor and related component fail, the failure occurred. In this approach, the reliability of sensor affects the whole system reliability.

By applying of this method, model of system is updated for each sensor placement scenarios and the occurrence probability will be calculated for top event (TE). According to the proposed probability, all scenarios can be prioritized.

Both the uncertainty and sensor reliability criterions are considered in this study. However, the prioritization of some scenarios is different in these two categories. One way to interpret of these results is to consider a weight factor for each criterion based on their importance in the optimization process. Considering sensors as high reliable components, a sensor is rarely failed. So the effect of sensor reliability must be less than prediction uncertainty in the optimization process. Determining a specific factor for each criterion is a complex process which depends on system functionality, environmental condition and quality of the sensor.

The other approach is applying field data, expert judgment and generic data for interpreting both proposed criterions together in an optimal process. This method is used when requirements of previous method are not available.

4 Case Study: Steam Turbine Monitoring and Optimal Sensor Placement

Steam turbines belong to a category of machines called turbo-machines. Main characteristic of turbo- machines is the energy conversion which takes place in a rotating wheel. The basic function of a steam turbine is to transform the thermal energy of steam into mechanical energy. The main components of a steam turbine are bearings, rotor, rotor blades, seals, diaphragms and casing.

According to the Sect. 2, in first step, components of steam turbine and their failure modes must be extracted. So the FMEA of a typical steam turbine is developed [27]. The simplified FMEA table is presented in Table 3.

Fault tree of the steam turbine is extracted as a functional model of system. Due to system complexity, the simplified form of its fault tree model is considered in Fig. 2.

Occurrence probabilities and monitoring costs for all failure modes are presented in Table 4. It should be noted that the values of proposed probabilities and costs are extracted through expert judgments and information in the literature [27].

In third step, the potential places of sensors are determined through SPI of each component. Considering simplified model, cutsets of the system are presented as below:

Turbine=Diaphragms+RotorBlades=S.Overheat+S.Humidity+Debris+Vibration+Crack

Based on Birnbaum Importance criterion, the importance of each component is calculated by proposed model. Finally, all failure modes are prioritized through proposed index as shown in Table 4. According to the proposed prioritization, 3 sensors are mounted on the system to monitor steam temperature, steam humidity and rotor vibration. In addition, performance of diaphragm and turbine are monitored by other independent sensors. Types of all applicable sensors are presented in Table 5.

The final model of system with potential places of sensors is presented in Fig. 3.

In the next step, state vector of the system is obtained. For calculating the occurrence probability of all system state vectors, it is necessary to know the prior occurrence probability of all failure modes. According to the literature and expert judgment, prior occurrence probability is provided for each failure mode. Based on existing standards, a criterion is specified for each failure mode. By utilization of Monte Carlo method, the occurrence probability is obtained for each state vector. Partial of proposed probabilities is presented in Table 6.

In the next step, information vector is extracted. To complete this task, it is necessary to determine the sensor placement scenarios in advance. According to the proposed model of steam turbine in Table 7, placement scenarios are considered as below:

To calculate the probability of each IV, probabilities of SVs which causing the occurrence of proposed IV, are added together. The result represents the occurrence

Table 3 Partial FMEA table of a typical steam turbine [27]

Component	Component function	Potential failure mode	Potential causes of failure	Occurrence (O)	Potential effects of failure	Severity (S)	Detection/Control	Detection (D)	RPN (O × S×D)
Diaphragm	Convert thermal energy to kinetic energy by accelerating the steam	Erosion	Penetration of solid particles from or droplets from steam	4	• High wear rate • Formation of crack • Breaking of nozzles • Vibration	4	Condition monitoring/visual inspection	3	48
		Scaling	Too dry steam	2	• Decreased pressure after turbine stage • Efficiency drop • Performance drop • Drop in mass flow • Vibration	3		3	18
		Corrosion	Exposure to the corrosive substance in the steam	1	Wear	2		3	6

(continued)

Table 3 (continued)

Component	Component function	Potential failure mode	Potential causes of failure	Occurrence (O)	Potential effects of failure	Severity (S)	Detection/Control	Detection (D)	RPN (O × S×D)
Rotor Blades	Convert kinetic energy to mechanical energy	Erosion	Penetration of solid particles from or droplets from steam	3	• High wear rate • Formation of crack • Breaking of blades • Vibration	5	Condition monitoring/visual inspection	3	45
		Cracking	• Fatigue • Vibration	1	Breaking of blades	5		4	20
		Scaling	Too dry steam	2	• Decreased pressure after turbine stage • Efficiency drop • Performance drop • Drop in mass flow • Vibration	3		2	12

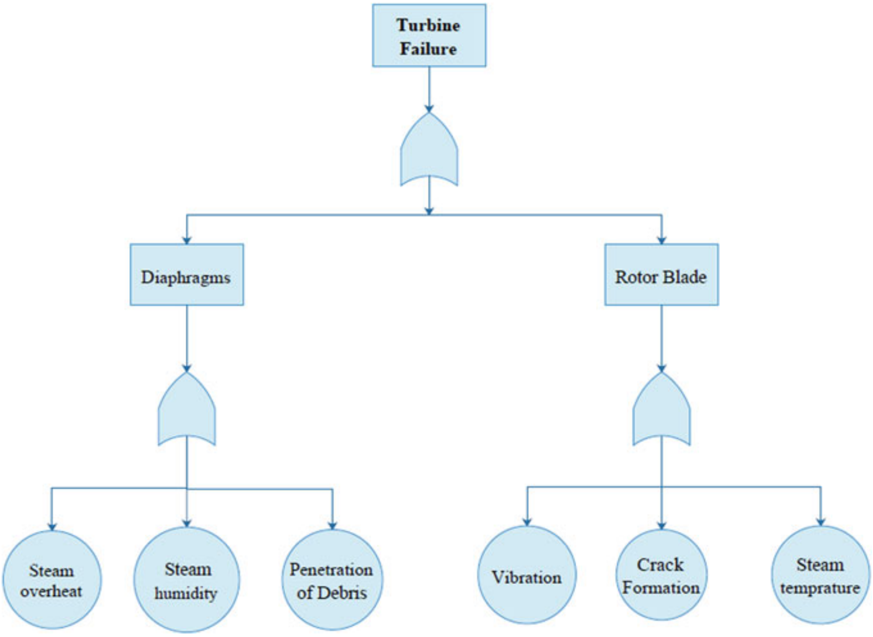


Fig. 2 Simplified fault tree of the steam turbine

Table 4 Importance index for all failure modes

Failure mode	Failure rate (per 10^6 h)	Monitoring cost	Importance index
Penetration of debris	57	Not Possible	–
Steam overheat	28	10 unit	0.1
Vibration and ageing	28	20 unit	0.05
Steam humidity	28	30 unit	0.03
Crack formation	14	100 unit	0.01

Table 5 Type of sensors

Sensor number	Sensor type
1	Tachometer
2	Wireless accelerometer
3	Accelerometer
4	Thermometer

probability of considered IV. As an example, IVs and their occurrence probabilities are presented for first scenario as shown in Table 8.

Extracting IVs and related probabilities, step 6 and step 7 are applied on the system. Then inverse of occurrence probabilities' variance for all failure modes has been calculated and added together for each scenario as it is shown in Table 9.

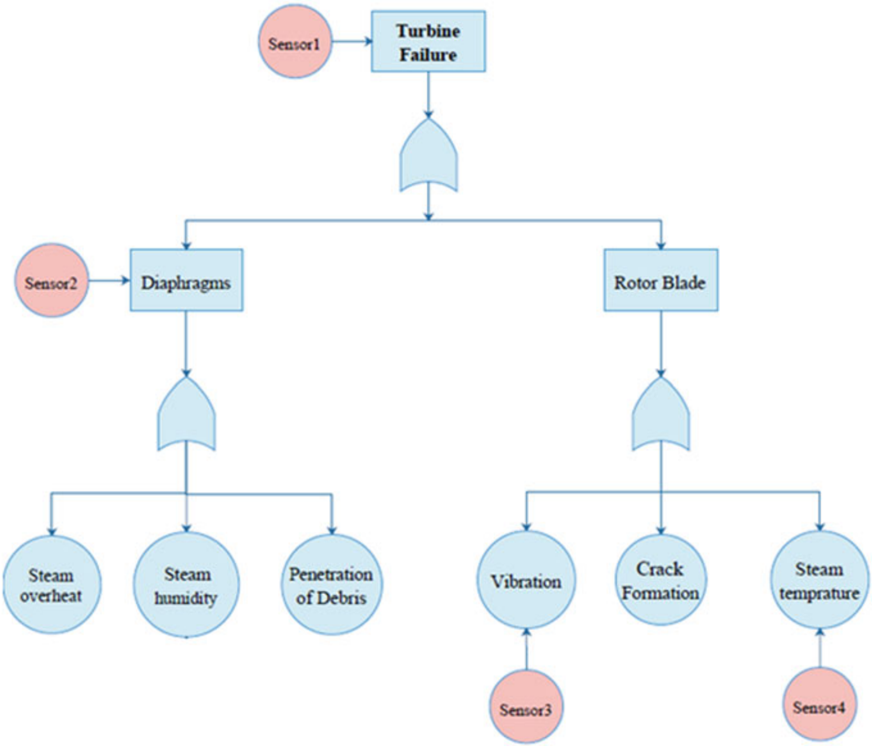


Fig. 3 Final Model of system with potential places for sensors

Table 6 State vectors of the steam turbine

Failure mode	SV1	SV2	SV3	SV4	SV5	SV6	…
Overheat	1	1	0	0	0	0	…
Humidity	1	0	1	0	0	0	…
Crack	1	0	0	1	0	0	…
Vibration	1	0	0	0	1	0	…
Debris	1	0	0	0	0	1	…
Occurrence probability	0.00001	0.0011	0.1509	0.0187	0.0011	0.0186	…

Table 7 Sensor placement scenarios

Scenario number	Sensor number
Scenario 1	Sensor1, Sensor2, Sensor3, Sensor4
Scenario 2	Sensor1, Sensor2, Sensor3
Scenario 3	Sensor1, Sensor2, Sensor4
Scenario 4	Sensor1, Sensor3, Sensor4
Scenario 5	Sensor2, Sensor3, Sensor4

Table 8 Occurrence probabilities of IVs

Sensor number	IV1	IV2	IV3	IV4	IV5	IV6	IV7
1	1	1	1	1	1	1	0
2	1	0	1	1	1	0	0
3	1	1	0	1	0	0	0
4	1	0	0	0	1	0	0
Probability	2.5e-6	0.001	0.1771	2.32e-4	0.0014	0.0186	0.8016

Table 9 Amount of uncertainty index for all scenarios

Scenario Number	Uncertainty Index (Σ 1/(variance of each component)
Scenario 1	1.12e+09
Scenario 2	6.9e+08
Scenario 3	4.26e+08
Scenario 4	7.83e+08
Scenario 5	3.9e+08

Table 10 Failure rate for different type of sensors [28]

Sensor Type	Typical failure rate (per 10^6 h)
Tachometer	80
Wireless accelerometer	1
Accelerometer	13
Thermometer	3.5

This index reflects the uncertainty of system state prediction. According to results in Table 10, uncertainty is smaller in state prediction of system in scenario 1. This result is obvious for the first scenario because all sensors are used in it. Among other 4 scenarios, fourth scenario is optimal since the system state prediction reports smaller uncertainty.

In the last step, sensors have been added to the system functional model as components. The modified model of system for first scenario is shown in Fig. 4.

Failure rate for different type of sensors are presented in Table 10 [28].

To calculate the occurrence probability of top event in each scenario, their minimal cutsets are extracted. For a specific time period, probabilities of top event have been calculated as Table 11.

According to Table 11, occurrence probability of top event in the first scenario is less than the others since all sensors have been used. For rest of 4 scenarios, scenario 2 and 3 has less top event probability so they are optimal scenarios based on the sensor reliability criterion.

As it can be seen, the prioritization of scenarios is different for any of uncertainty and sensor reliability criterion. This indicates that not only the uncertainty affects the optimization process of sensor placement, but also the sensor reliability is

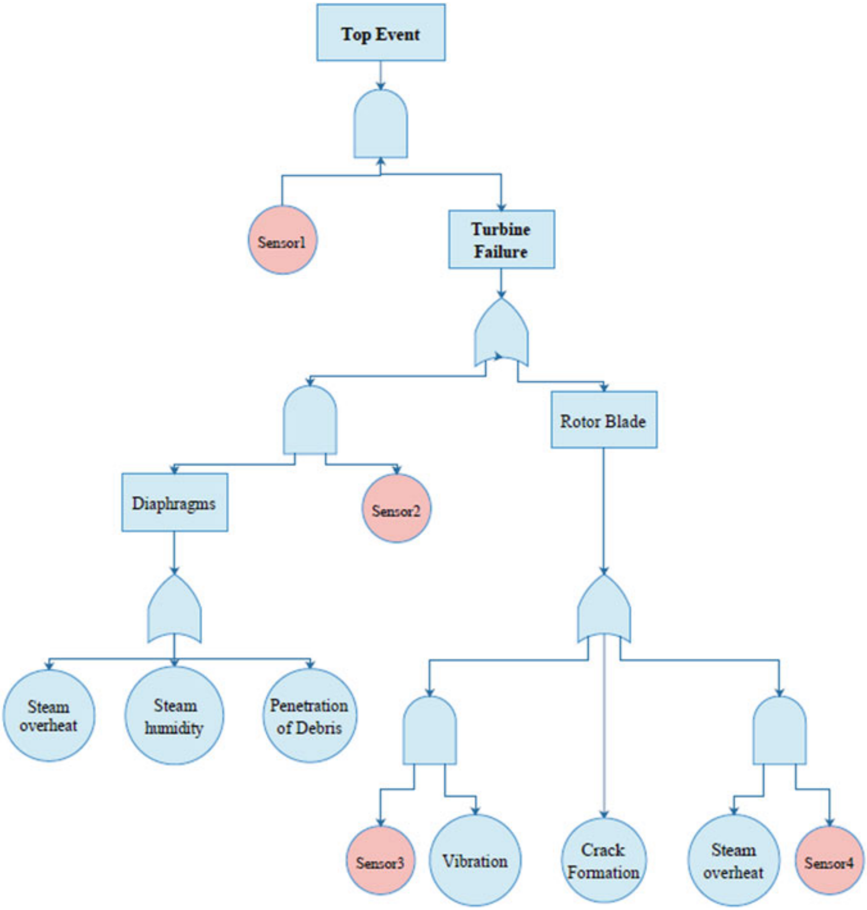


Fig. 4 Modified model of the first scenario

Table 11 Occurrence probability of TE for all scenarios

Scenario	Cutsets	Occurrence probability of T.E
1	S1.S3.V + S1.C + S1.S4.O + S1.S2.O + S1.S2.H + S1.S2.PD	2.02e-6
2	S1.S3.V + S1.C + S1.O + S1.S2.H + S1.S2.PD	6.013e-6
3	S1.V + S1.C + S1.S4.O + S1.S2.O + S1.S2.H + S1.S2.PD	6.017e-6
4	S1.S3.V + S1.C + S1.O + S1.H + S1.PD	1.5e-3
5	S3.V + C + S4.O + S2.O + S2.H + S2.PD	0.0221

*V = vibration, C = crack, O = overheat, H = humidity, PD = penetration of debris considered as well. The importance of each criterion is determined based on system functionality, complexity and expert judgment. Final prioritization of scenarios is developed by considering both criterions together.

5 Discussion and Conclusion

The accuracy of system state prediction methods (e.g., condition monitoring) is strongly depends on sensor network arrangement. Therefore optimization of proposed arrangement increases the focus on the important components of the system in order to reduce the maintenance costs. Based on the results of this study, the main effective factors include uncertainty of sensors' information and the reliability of the sensor itself on optimal sensor placement.

Studying both uncertainty and sensor reliability indexes separately, it is observed that the optimal sensor placement offer different prioritizations of sensor placement scenarios based on proposed two criterions. Therefore it is necessary to select the more important one or applying both of them criterions in optimization process.

In concluding, sensors are high-reliable components and the time to failure (TTF) of them is much more than TTF of common mechanical components. So in a system with high-reliable components, TTF of components are close to TTF of sensors. So the reliability of sensors is determinant in such a system in the case of condition monitoring. On the other hand, in a system with common components, because of wide gap between TTF of components versus sensors, the effect of sensor reliability is negligible on optimal sensor placement. As a result, it is necessary to study the system functionality and reliability of its components before making any decision about the optimal sensor placement.

References

1. Moubray J (2004) Reliability-centered maintenance, 2nd edn. Elsevier, Oxford
2. Randall RB (2011) Vibration-based condition monitoring, 1st edn. Wiley, New Delhi
3. Beebe RS (2004) Predictive maintenance of pumps using condition monitoring, 1st edn. Elsevier, Oxford
4. Van Horenbeek A, Van Ostaeyen J, Duflou JR, Pintelon L (2013) Quantifying the added value of an imperfectly performing condition monitoring system—application to a wind turbine gearbox. Reliab Eng Syst Saf 111:45–57
5. Zhang G, Vachtsevanos G (2007) A methodology for optimum sensor localization/selection in fault diagnosis. In: IEEE Aerospace conference, Montana
6. Guratzsch RF, Mahadevan S (2007) Structural health monitoring sensor placement optimization under uncertainty. American Institute of Aeronautics and Astronautics
7. Wang L, Gao RX (2006) Condition monitoring and control for intelligent manufacturing, 1st edn. Springer, London
8. Kammer DC (1991) Sensor placement for on-orbit modal identification and correlation of large space structures. J Guid Control Dyn 14(2):251–259

9. Stephan C (2012) Sensor placement for modal identification. Mech Syst Signal Process 27:461–470
10. Yao L, Sethares WA, Kammer DC (1993) Sensor placement for on-orbit modal identification via a genetic algorithm. AIAA J 31(10):1922–1928
11. Wang C, Gao RX (2000) Sensor placement strategy for in-situ bearing defect detection. In: Proceedings of the IEEE instrumentation and measurement technology conference, Baltimore
12. Kammer DC, Tinker Michael L (2004) Optimal placement of triaxial accelerometers for modal vibration tests. Mech Syst Signal Process 18:29–41
13. Camelio JA, Jack HS, Hyunjune Y (2005) Sensor placement for effective diagnosis of multiple faults in fixturing of compliant parts. J Manuf Sci Eng, Trans ASME 127:68–74
14. Kammer DC (2005) Sensor set expansion for modal vibration testing. Mech Syst Signal Process 19:700–713
15. Alkhadafe H, Al-Habaibeh A, Daihzong S, Lotfi A (2012) Optimising sensor location for an enhanced gearbox condition monitoring system. In: 25th International congress on condition monitoring and diagnostic engineering, Huddersfield, UK
16. Meo M, Zumpano G (2005) On the optimal sensor placement techniques for a bridge structure. Eng Struct 27:1488–1497
17. Rosich A, Sarrate R, Nejjari F (2003) Optimal sensor placement for FDI using binary integer linear programming. Berlin declaration on open access to knowledge in the sciences and humanities
18. Sarrate R, Puig V, Escobet T, Rosich A (2007) Optimal sensor placement for model-based fault detection and isolation. In: Proceedings of the 46th IEEE conference on decision and control, New Orleans, LA, USA
19. Bhushan M, Narasimhan S, Rengaswamy R (2003) Sensor network reallocation and upgrade for efficient fault diagnosis. In: Fourth international conference on foundations of computer-aided process operations, Coral Springs, Florida, USA
20. Bhushan M, Rengaswamy R (2002) Comprehensive design of a sensor network for chemical plants based on various diagnosability and reliability criteria. 2. Applications. Ind Eng Chem Res 41:1840–1860
21. Assaf T, Dugan JB (2008) Diagnosis based on reliability analysis using monitors and sensors. Reliab Eng Syst Saf 93:509–521
22. Duan R, Ou D, Dong D, Zhou H (2011) Optimal sensor placement for fault diagnosis based on diagnosis cost specifications. J Comput Inf Syst 7(9):3253–3260
23. Flynn Eric B, Todd Michael D (2010) A Bayesian approach to optimal sensor placement for structural health monitoring with application to active sensing. Mech Syst Signal Process 24:891–903
24. Jackson C, Mosleh A (2012) Bayesian inference with overlapping data for systems with continuous life metrics. Reliab Eng Syst Saf 106:217–231
25. Pourali M, Mosleh A (2012) A Bayesian approach to functional sensor placement optimization for system health monitoring. In: IEEE conference on prognostics and health management, Denver Co USA
26. Modarres M, Kaminskiy M, Krivtsov V (1999) Reliability engineering and risk analysis: a practical guide. Marcel Dekker Inc, NewYork
27. Gunnarsson A (2013) Maintenance of the steam turbines at Hellisheidi power plant. Faculty of Industrial Engineering, Mechanical Engineering and computer science, School of Engineering and Natural Science, University of Iceland, Iceland
28. Jones TL (2010) Handbook of reliability prediction procedures for mechanical equipment. Naval Surface Warfare Center, Carderock Division, West Bethesda, Maryland

Estimation of the Reliability of Rolling Element Bearings Using a Synthetic Failure Rate

Urko Leturiondo, Oscar Salgado and Diego Galar

Abstract As rolling element bearings are key parts of rotating machinery, the estimation of their reliability is very important. In this context, different standards and research articles propose how to estimate fatigue life for different levels of reliability. However, when trying to do calculations based on data from a real system, there are many difficulties because of economic and safety reasons. Consequently, the use of physical models to simulate the cases that are difficult to reproduce in a real system allows us to generate synthetic data related to them. Thus, in this paper a synthetic failure rate of rolling element bearings is calculated using a physical modelling approach. A multi-body model of a bearing is used in order to obtain its dynamic response in non-stationary conditions and in different degradation levels. Thus, synthetic data are generated to cover a range of degradation related to geometric changes in the surface of the parts of the bearing. Some of the output variables of these synthetic data, such as vibration, are used as covariates of a proportional hazard model, which is then trained to estimate the reliability of the bearing. In this way, a synthetic failure rate is obtained in such a way that it can improve the failure rate given by the manufacturers

Keywords Rolling element bearing · Synthetic data · Proportional hazard model · Reliability

U. Leturiondo · O. Salgado
Mechanical Engineering, IK4-Ikerlan, J. M. Arizmendiarrieta 2,
20500 Arrasate-Mondragón, Gipuzkoa, Spain
e-mail: osalgado@ikerlan.es

U. Leturiondo (✉) · D. Galar
Division of Operation and Maintenance Engineering, Luleå University of Technology,
971 87 Luleå, Sweden
e-mail: uleturiondo@ikerlan.es

D. Galar
e-mail: diego.galar@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_8

1 Introduction

Rolling element bearings are widely used parts in rotating machinery. Thus, they are very important components in different sectors such as railways or wind energy, among others. That is the reason why the reliability of rolling element bearings is a key issue for the correct operation of the systems in which they are placed.

There are many failure modes that can affect to that operation [6], e.g. fatigue, corrosion, electrical erosion, etc. According to Ferreira et al. [2], fatigue is the main reason for the appearance of defects in the different parts of rolling element bearings. This failure mode has been broadly studied by the manufacturers and the selection of the bearings is commonly done based on the knowledge behind the fatigue development. Thus, the international standard ISO 281 [5] is the main reference in this field. It gives the relation between the reliability of a rolling element bearing and its life for a given stationary operating condition.

If the relation between reliability and life is wanted to be obtained by using data acquired from a real system instead of using the formulae proposed by ISO 281 [5], there are some limitations. Those are related to the fact that some operating conditions, specially related to damaged states or extreme conditions, cannot be reproduced due to safety reasons and because of the economic cost related to both the development of the testing and the likelihood of happening serious consequences in the system.

Thus, the use of data obtained by simulations carried out by using physical models is an alternative to real data. These data are called synthetic data [12] and they can be generated in such a way that they give information about the performance of a system in the aforementioned operations which are difficult to reproduce.

In this paper a methodology for obtaining a synthetic failure rate using synthetic data is presented. A physical model is used to generate synthetic data of the dynamics of rolling element bearings in different degradation levels. Once the degradation level is associated with its corresponding value of reliability by the use of a degradation curve and a reliability-life curve, the synthetic data are taken as input data to fit a proportional hazard model.

Proportional hazard models are statistical models that can be used for the estimation of the hazard of systems by means of the use of influential factors. Specifically, indicators from a time-domain analysis of the velocity of the inner ring are used to train the model. The analysis of the fitness of the model gives the key to know which of the indicators provides more meaningful information about the reliability of the simulated bearing.

This strategy will help to have estimations of the reliability of rolling element bearings that are able to adapt to the context in which they are operating [8]. In this way the reliability of a machine that suffers highly varying operating conditions can be accurately obtained and, therefore, appropriate maintenance actions can be taken by using this information.

This paper is structured as follows: the methodology followed in order to obtain the synthetic failure rate is explained in Sect. 2; Sect. 3 shows the results obtained by the application of the methodology to a specific bearing; finally, the conclusions are presented in Sect. 4.

2 Through the Synthetic Failure Rate

In this section the methodology that has been developed in order to estimate the reliability of rolling element bearings using a synthetic failure rate is presented. First, how the synthetic data is generated is explained; then, the degradation and reliability curves that have been used in this study are presented; finally, the calculation of the hazard model is shown.

2.1 Synthetic Data Generation

As stated in Sect. 1, synthetic data can be generated in different operating conditions and considering different degradation levels using physical models. In the field of rolling element bearings, there are many models that reproduce their response, taking into account different features of the physics behind these components.

In this work the multi-body model developed by Leturiondo et al. [10] is used to carry out the simulations needed to generate synthetic data. This model is able to simulate the dynamics of any kind of rolling element bearing in any configuration, considering each element of the bearing as a rigid body with 6 degrees of freedom. The metal-metal contacts between those elements are modelled using the Hertz contact and elastohydrodynamic lubrication theories. Besides that, local defects are modelled as geometric changes in the surface of the elements.

The bearing selected to be simulated is a single-row deep-groove ball bearing with 8 balls; its dimensions are shown in Table 1.

All components are considered to be made of steel with the following properties: modulus of elasticity of 207 GPa, Poisson number of 0.3 and density of 7830 kg/m^3. A constant value of 30 °C is chosen as the operating temperature. The

Table 1 Dimensions of the simulated bearing

Dimension	Value (mm)
Ball diameter, D_w	22.46
Outer raceway diameter, d_o	87.73
Inner raceway diameter, d_i	42.79
Pitch diameter, D_{pw}	65.26
Outer groove radius, r_o	11.6792
Inner groove radius, r_i	11.6792

Table 2 Defect area in the defined 16 degradation levels

Code	Area (mm^2)
A_1	0.390625
A_2	0.78125
A_3	1.171875
A_4	1.5625
A_5	1.953125
A_6	2.34375
A_7	2.734375
A_8	3.125
A_9	3.515625
A_{10}	3.90625
A_{11}	4.296875
A_{12}	4.6875
A_{13}	5.078125
A_{14}	5.46875
A_{15}	5.859375
A_{16}	6.25

values of the dynamic viscosity η_0 and the viscosity-pressure coefficient α at this temperature are 0.04 Pa s and $1.2 \cdot 10^{-8}$ Pa^{-1}, respectively.

The inner ring is selected as the rotating one; the outer ring is assumed to be located in a rigid housing. Regarding the operating conditions, a constant value of 20 rad/s for the inner ring speed and a constant value of 300 N for the radial load applied in the same ring are selected. It should be noted that the load is applied to the bearing in vertical direction.

Different simulations have been carried out in order to obtain synthetic data in different degradation levels. This degradation is modelled as a size variant local geometrical change in the most loaded zone of the outer raceway of the rolling element bearing. Thus, 16 simulations have been done by taking a value of the spall areas A_1 to A_{16} shown in Table 2 for each simulation. It should be noted that the defect size A_{16} is equal to that at which the industry considers a rolling element bearing to have reached its faulty state [1].

These 16 simulations have been carried out using the software Dymola®, studying the response of the rolling element bearing during 5 s. The values used for the time sampling period and for the integration tolerance are 1 ms and 10^{-4}, respectively.

The model gives information regarding different physics of the bearing. Thus, the linear and angular position of each element of the bearing (rings, rolling elements and cage) can be obtained, as well as other variables related to the contact between the elements, such as the contact loads (both normal and tangential) and the lubricant film thickness. In this work the vibratory response of the rolling element bearing is used as the input for training the hazard model, in particular, the vertical velocity of the inner ring due to its observability.

2.2 Degradation and Reliability Curves

The relation between the degradation of a bearing and the time in which a specific level of degradation occurs is necessary in order to calculate the bearing life. There are many theories regarding this issue in the literature, from simple degradation models to others with a higher complexity. In this paper the Paris' law has been selected for the degradation variation and, thus, the prediction proposed by Li et al. [11] is used as an approximation to obtain the bearing running time at which the spall areas A_1 to A_{16} occur. The results for this degradation-time relation are shown in Fig. 1.

As stated in Sect. 2.1, the faulty limit is reached when the defect size is equal to 6.25 mm^2. This defect size is obtained when a bearing operates during a time t_f equal to $12.3 \cdot 10^6$ revolutions. Therefore, it can be assumed that the life of a bearing in the degradation levels defined by the spall areas A_1 to A_{16} is equal to the difference between t_f and the time in which the aforementioned spall areas occur.

Regarding the reliability, the curve proposed by the ISO 281 [5] has been selected. First of all, the basic rating life L_{10} is calculated using the properties of the configuration of the bearing, the geometrical properties shown in Table 1 and the data regarding the loading conditions. Then, the modified rating life L_{nm} is calculated by multiplying L_{10} by the different life modification factors a_1 presented in the standard for values of the reliability from 95 to 99.95 %. This curve is modified randomly in order to represent the variations that the reliability of a bearing can suffer due to the fact that the data of the response of the bearing are generated synthetically. The relation between the reliability and the life of the bearing is shown in Fig. 2.

Following this approach, the values of life and reliability for each degradation level shown in Table 3 are obtained. It can be seen that the value of the reliability for the last level of degradation is very high whereas its value for life is very low, which means that the defect produces a situation in which the bearing is near to fail.

Fig. 1 Relation between degradation and the operation time

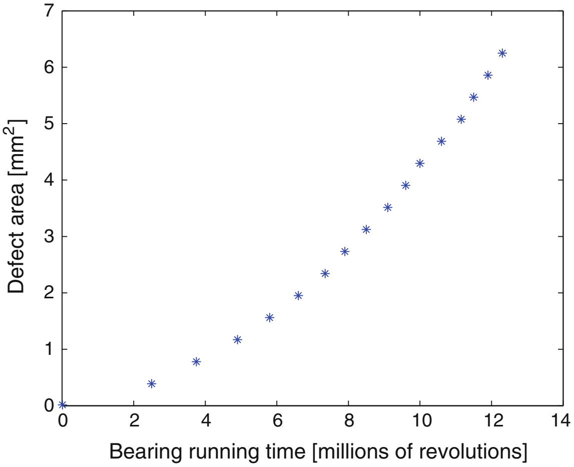


Fig. 2 Relation between the life of the bearing and its reliability

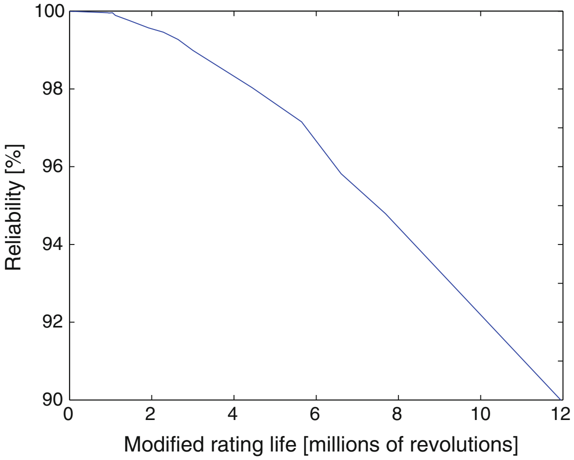


Table 3 Input data for the proportional hazard model

Code	Life (rev.·10^6)	Reliability (%)
A_1	9.83	92.38
A_2	8.58	93.78
A_3	7.43	95.03
A_4	6.53	95.92
A_5	5.73	97.03
A_6	4.98	97.64
A_7	4.43	98.04
A_8	3.83	98.44
A_9	3.23	98.84
A_{10}	2.73	99.2
A_{11}	2.33	99.43
A_{12}	1.73	99.65
A_{13}	1.18	99.87
A_{14}	0.83	99.96
A_{15}	0.43	99.98
A_{16}	0.03	99.99

2.3 Hazard Models for Reliability Estimation

Nowadays the importance of the determination of the condition of a system has raised for its use in diagnosis and prognosis processes. Thus, the decisions of the actions that have to be carried out for maintenance are easier to take. If it is properly done, this entails a reduction of machinery downtime and the inventory of spares, which has a direct relation with the decrease of the risk of having a failure and, finally, with the reduction of the costs related to maintenance [4].

For this purpose, the estimation of the reliability of the assets related to its remaining useful life is a key. It should be taken into account that there are many factors of the operation of the systems that have a great influence in this estimation. There are many statistical models to obtain the relation between these factors, called covariates, and the hazard of an asset. These models have been especially used in the fields of reliability and biomedicine, being proportional hazard models the origin of most of them [3].

A proportional hazard model consists in a function formed by the product of a baseline hazard rate and a positive function described by covariates that have a multiplicative effect on the baseline hazard rate and a regression parameter for each of these covariates. Thus, it is expressed as:

$$h(t, z) = h_0(t) \cdot \psi(\boldsymbol{\beta}^T z) \tag{1}$$

where $h(t,z)$ is the hazard rate, $h_0(t)$ is the baseline hazard rate, t is the time, z is the vector of covariates, $\boldsymbol{\beta}$ is the vector of regression parameters and ψ is the positive function, being the exponential the most used one.

Proportional hazard models are very useful when only the final remaining useful life of the system and its confidence limit are required, and when there are data available for the failure modes being modelled [14]. Besides that, the aforementioned multiplicative effect is a realistic and reasonable assumption for the relation between covariates and the hazard rate.

The review of Gorjian et al. [3] shows different methods existing in literature based on proportional hazard models, classified by the fact of them being non-parametric or semi-parametric. The methods shown in the following list are some of the ones found in this first group:

- Stratified proportional hazard model: it considers a population divided in a number N of levels (for example, different operating conditions). Thus, there are N baseline hazard functions, each defining the distinctive features of each level, whereas the regression coefficients are the same for all the levels. The expression of this kind of model for the jth level is the following:

$$h_j(t, z) = h_{0j}(t) \cdot \exp(\boldsymbol{\beta}^T z) \tag{2}$$

- Two-step regression model: it assumes that there is a difference in the effect of covariates during time, in such a way that a breakpoint is defined at time B. Thus, the model before B has time-dependent regression parameters $\alpha_i = \beta_i(t) = \beta_i \cdot \exp(-\gamma_i \cdot t)$, whereas after B regression parameters are constant. Equation 3 shows the formula for this kind of model. The main limitations for this approach are the difficulty to estimate the breakpoint and the assumption of all the covariates having the same breakpoint

$$h(t,z) = \begin{cases} h_0(t) \cdot \exp(\boldsymbol{\alpha}^T z) & t \le B \\ h_1(t) \cdot \exp(\gamma^T z) & t > B \end{cases} \tag{3}$$

- Additive hazard model: this strategy changes one of the basic features of the original proportional hazard model: the multiplicative effect of the covariates. This is changed to take an additive effect, as expressed in Eq. 4. It gives good results regarding the effect of repairs but it can only be used to model those failure modes that imply a jump ψ in the hazard.

$$h(t,z) = h_0(t) + \psi(\boldsymbol{\beta}^T z) \tag{4}$$

- Mixed model: it takes both the additive and the multiplicative approaches to take advantage of the benefits of each method. It is formulated as:

$$h(t|z) = g\{\boldsymbol{\beta}_0 \boldsymbol{w}(t)\} + h_0(t) \cdot f\left\{\gamma_0^T \boldsymbol{y}(t)\right\} \tag{5}$$

  It can be seen that the covariates are separated in two groups ($\boldsymbol{w}$ and $\boldsymbol{y}$), having each of the covariate vector its corresponding regression parameter vector ($\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively) and link function (g and f, respectively). The mixed model fits better to the data rather than the classical model but has an extremely limited testing.
- Accelerated failure time model: accelerated tests are commonly done in industry in order to obtain the results of reliability, failure rate and life of different component and systems in a shorter period of time. Thus, this kind of model links the classical proportional hazard model with the accelerated reliability by means of multiplying the operation time by the effect of the covariates and using this value to determine the timely value of the baseline hazard rate, as it is expressed in the following equation:

$$h(t,z) = h_0(t \cdot \psi(\gamma^T z)) \cdot \psi(\gamma^T z) \tag{6}$$

The semi-parametric models take advantage of other models such as the Weibull distribution or logistic regression models in order to obtain more complex forms. The main drawback of this kind of models is the need to calculate more parameters than in the other models.

As stated before, these techniques have been widely use in different areas regarding reliability and biomedicine. An example of the application of proportional hazard models can be found in the work of Krivtsov et al. [9]. A survival analysis is carried out regarding the tread and belt separation of automobile tires. For that purpose, different tire design characteristics such as the tire age, the wedge gauge, the peel force or the percent of carbon black, among others, are used to fit a proportional hazard model. The analysis of the p-values for each regression coefficient gives an indication of what covariates are the most significant factors for the studied failure mode.

In this work, the original proportional hazard model is used as a first approximation to fit the synthetic data obtained from the physical model and the values of the reliability calculated by means of the degradation and reliability curves.

3 Results and Discussion

As stated in Sect. 2.1, the vibratory response of the rolling element bearing is used to estimate its reliability. Figure 3 shows the vertical velocity of the inner ring.

In order to obtain the inputs for the proportional hazard model, some indicators are extracted from the velocity signal. For this purpose, time-domain analysis techniques are applied. Specifically, features that have been widely used in the field of diagnosis of rolling element bearings and transmissions are extracted from the synthetic data [7, 13]. Thus, the following indicators have been calculated:

- Mean:

$$\mu = \frac{\Delta t}{t_1 - t_0} \cdot \sum_{t=t_0}^{t_1} x(t) \tag{7}$$

- Standard deviation:

$$\sigma = \sqrt{\frac{\Delta t}{t_1 - t_0} \cdot \sum_{t=t_0}^{t_1} |x(t) - \mu|^2} \tag{8}$$

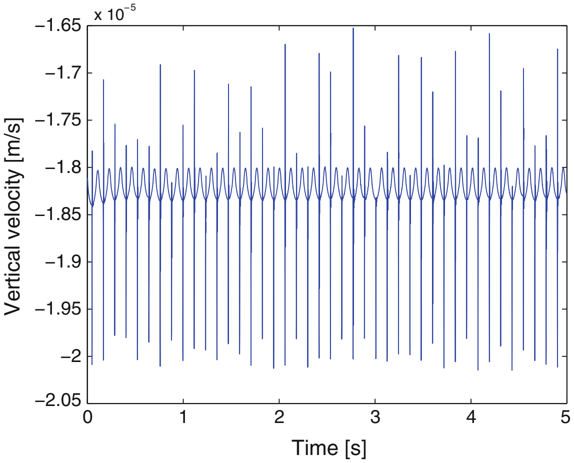


Fig. 3 Vertical velocity of the inner ring obtained by the physical model

- Skewness:

$$\gamma = \frac{\frac{\Delta t}{t_1 - t_0} \cdot \sum_{t=t_0}^{t_1} |x(t) - \mu|^3}{\sigma^3} \tag{9}$$

- Kurtosis:

$$\kappa = \frac{\frac{\Delta t}{t_1 - t_0} \cdot \sum_{t=t_0}^{t_1} |x(t) - \mu|^4}{\sigma^4} \tag{10}$$

- Peak:

$$x_k = \max|x(t)| \tag{11}$$

- Root mean square:

$$x_{RMS} = \sqrt{\frac{\Delta t}{t_1 - t_0} \cdot \sum_{t=t_0}^{t_1} |x(t)|^2} \tag{12}$$

- Crest factor:

$$CF = \frac{x_k}{x_{RMS}} \tag{13}$$

- Shape factor:

$$SF = \frac{x_{RMS}}{\frac{\Delta t}{t_1 - t_0} \cdot \sum_{t=t_0}^{t_1} |x(t)|} \tag{14}$$

- Impact factor:

$$IF = \frac{x_k}{\frac{\Delta t}{t_1 - t_0} \cdot \sum_{t=t_0}^{t_1} |x(t)|} \tag{15}$$

- Energy operator (EO): calculated as the kurtosis of the following signal:

$$\left[x(t_p)\right]^2 - \left[x(t_p - \varDelta t) \cdot x(t_p + \varDelta t)\right], \qquad t_0 \le t_p \le t_1 \tag{16}$$

where $x(t)$ is the velocity signal, t_0 is the initial time (0 s), t_1 is the final time (5 s) and $\varDelta t$ is the time sampling period (1 ms).

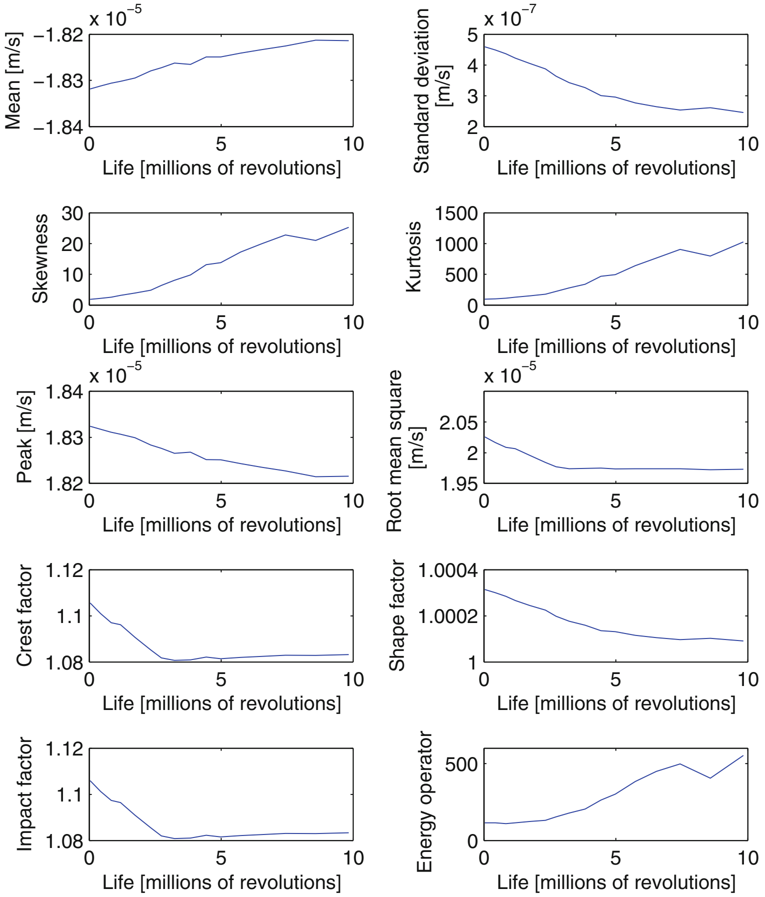


Fig. 4 Relation between the indicators and the bearing life

Thus, one feature is calculated for each vibration signal in the different degradation levels. This leads to 10 values for each of the 16 damaged scenarios. At the end of the day, one or a combination of indicators can be used as the covariate of the proportional hazard model. With the objective of better understanding the reliability changes that occur in rolling element bearings, the selection of appropriate indicators is crucial.

Figure 4 shows the relation between the ten indicators used in this analysis and the data of bearing life presented in Table 3. It can be seen that in general there is a relation between these values. It should be noted that some indicators provide a clear relationship only for the last stages of the degradation, as it occurs in the case of the crest factor.

Table 4 Results of the proportional hazard models taking each indicator as an only covariate for each model

Covariate	β	Stand. error	p-value
μ	$4.84 \cdot 10^6$	$1.21 \cdot 10^7$	0.6902
σ	$-1.85 \cdot 10^8$	$6.75 \cdot 10^7$	0.0062
γ	1.81	0.72	0.0126
κ	$3.7 \cdot 10^{-2}$	$1.45 \cdot 10^{-2}$	0.0107
x_k	$-4.84 \cdot 10^6$	$1.16 \cdot 10^7$	0.6774
x_{RMS}	$-4.46 \cdot 10^6$	$2.3 \cdot 10^6$	0.0524
CF	−108.99	52.09	0.0364
SF	−108.99	$4.89 \cdot 10^3$	0.9822
IF	−108.99	51.91	0.0358
EO	$4.84 \cdot 10^{-2}$	$1.71 \cdot 10^2$	0.0048

The values of the features presented in Fig. 4 are used as the covariates of the proportional hazard model (i.e. predictor values), and the values of the reliability, which correspond to the hazard rate, are used as the objective values for the function that aims to be fitted. In this case, the baseline hazard model is equal to the nominal reliability-life curve given by the international standard ISO 281 [5].

In order to calculate the regression parameters $\boldsymbol{\beta}$ the *coxphfit* function of Matlab® has been used. First of all, an analysis is done taking each covariate to train a proportional hazard model. Thus, 10 models are obtained, each of them having an only regression parameter β. The results of this analysis are shown in Table 4.

Having a p-value lower than 0.05 has been taken as the criterion to select those covariates that are statistically significant. Thus, the mean, the peak value, the root mean square and the shape factor can be excluded from the study.

Once the significant covariates are identified, their combination by pairs is done. It should be highlighted that some of the couples have not been analysed due to the correlation between the indicators in their statistical definition. Thus, and taking a look to Fig. 4, the combination between the skewness and the kurtosis as well as the combination between the crest factor and the impact factor are rejected.

The results for the proportional hazard models fit by couples of covariates are shown in Table 5. There is not any couple that fits the significance criterion of having a p-value less than 0.05. Consequently, any of the pairs defined is able to give an accurate estimation of the reliability.

Thus, the use of some of the covariates individually is the only way to construct a proportional hazard model using the indicators presented in this section. Specifically, the values of the standard deviation, the skewness, the kurtosis, the crest factor, the impact factor and the energy operator are the ones that can be used to reproduce the reliability of a rolling element bearing in the defined conditions.

Table 5 Results of the proportional hazard models taking the indicators by pairs for each model

Covariates	p-value (1st covariate)	p-value (2nd covariate)
σ and γ	1	1
σ and κ	0.9658	0.0915
σ and *CF*	0.9498	0.0926
σ and *IF*	0.9976	0.9974
σ and *EO*	0.958	0.1575
γ and *CF*	0.1178	0.8450
γ and *IF*	1	1
γ and *EO*	0.3397	0.0865
κ and *CF*	0.832	1
κ and *IF*	0.0954	0.9648
κ and *EO*	0.0615	0.6636
CF and EO	0.0417	0.5150
IF and *EO*	0.9549	0.1607

4 Conclusions

The estimation of the reliability of an asset in general and the reliability of a rolling element bearing in particular is essential in order to carry out an optimum asset health management and minimize maintenance costs. As the collection of data from real systems for this estimation is difficult or even impossible to do in certain conditions, the use of synthetic data generated by physical models gains importance. Thus, the outputs of simulations can be used as inputs for a statistical model, obtaining a synthetic failure rate.

In this research work, a multi-body model of rolling element bearings is used to generate data related to different degradation levels. Time-domain analysis is carried out in order to obtain statistical indicators of the vertical velocity signal of the inner ring of the bearing. In order to obtain the relationship between the degradation levels of the synthetic data and their corresponding reliability two degradation models have been used. Then, the indicators are used to fit a proportional hazard model, which is a kind of statistical tool commonly used for reliability purposes.

Results show that some of the indicators that have been used represent properly the reliability of a bearing. In particular, the standard deviation, the skewness, the kurtosis, the crest factor, the impact factor and the energy operator give good results. However, their combination by pairs has failed to obtain good predictions.

Future work asks for determining other features that can fit better the reliability curve. Besides that, the extension of this methodology to cover all the operating conditions of a system can be useful to monitor the system and take advantage of proportional hazard models with maintenance purposes.

At the end of the day, there is a real need for customizing the presented methodology to specific applications, doing the fitting of the models with synthetic data generated by physical models as well as data provided by the manufacturers of the components. Thus, the tuning of the model can be done by the combination of both data, personalising the model to the requirements of the industry.

Acknowledgments This study is partially funded by the Basque Government under the Emaitek 2014 Program and by the Ministry of Economy and Competitiveness of the Spanish Government under the Retos-Colaboración Program (LEMA project, RTC-2014-1768-4). Any opinions, findings and conclusions expressed in this article are those of the authors and do not necessarily reflect the views of funding agencies. The authors would also like to thank Fundación de Centros Tecnológicos - Iñaki Goenaga.

References

1. Camci F, Medjaher K, Zerhouni N, Nectoux P (2013) Feature evaluation for effective bearing prognostics. Qual Reliab Eng Int 29(4):477–486
2. Ferreira JLA, Balthazar JC, Araujo APN (2003) An investigation of rail bearing reliability under real conditions of use. Eng Fail Anal 10(6):745–758
3. Gorjian N, Ma L, Mittinty M, Yarlagadda P, Sun Y (2009) A review on reliability models with covariates. In: Proceedings of the 4th world congress on engineering asset management
4. Heng A, Zhang S, Tan ACC, Mathew J (2009) Rotating machinery prognostics: State of the art, challenges and opportunities. Mech Syst Signal Process 23(2):724–739
5. ISO (2007) 281: Rolling bearings—dynamic load ratings and rating life. Geneva, Switzerland
6. ISO (2004) 15243: Rolling bearings—damage and failures—terms, characteristics and causes. Geneva, Switzerland
7. Jin X, Zhao M, Chow TWS, Pecht M (2014) Motor bearing fault diagnosis using trace ratio linear discriminant analysis. IEEE Trans Industr Electron 61(5):2441–2451
8. Johansson C-A, Simon V, Galar D (2014) Context driven remaining useful life estimation. Procedia CIRP 22:181–185
9. Krivtsov VV, Tananko DE, Davis TP (2002) Regression approach to tire reliability analysis. Reliab Eng Syst Saf 78(3):267–273
10. Leturiondo U, Mishra M, Salgado O, Galar D (2014) Nonlinear response of rolling element bearings with local defects. In: Proceedings of the eleventh international conference on condition monitoring and machinery failure prevention technologies
11. Li Y, Billington S, Zhang C, Kurfess T, Danyluk S, Liang S (1999) Dynamic prognostic prediction of defect propagation on rolling element bearings. Tribol Trans 42(2):385–392
12. Mishra M, Leturiondo-Zubizarreta U, Salgado-Picon O, Galar-Pascual D (2015) Modelización híbrida para el diagnóstico y pronóstico de fallos en el sector del transporte. Datos adquiridos y datos sintéticos, Dyna, 90(2):139–145
13. Samuel PD, Pines DJ (2005) A review of vibration-based techniques for helicopter transmission diagnostics. J Sound Vib 282(1–2):475–508
14. Sikorska JZ, Hodkiewicz M, Ma L (2011) Prognostic modelling options for remaining useful life estimation by industry. Mech Syst Signal Process 25(5):1803–1836

Nonlinear Process Monitoring Using Genetic Algorithms

Tawfik Najeh, Achraf Jaber Telmoudi and Lotfi Nabli

Abstract This paper suggests a new approach for fault detection using Genetic Algorithms (GAs). GAs are used to find the principal curve that summarize the data. The principal curve is a generation of linear Principal Component Analysis (PCA). Introduced by Hastie as a parametric curve, the original definition is based on the self-consistency property. The Hastie's theory encloses weaknesses in case of complex data structures or data with intersections. The existing principal curves methods employ the first component of the data as an initial estimation of principal curve that passes satisfactorily through the middle of data. However the needing of an initial line is the major inconvenient of this approach. In this work, we extend this problem in two ways. First, we introduce a new method based on GAs to find the principal curve. Second, potential application of principal curves in fault detection is proposed. An example is presented to prove the efficiency of the proposed algorithm to fault detection of nonlinear process.

Keywords Principal curves · Genetic algorithms · Fault detection

1 Introduction

The current trend in the industrial automation and industrial equipment leads to mechatronic systems ever more complex, working in an uncertain, changeable environment, corresponding to a permanent search for improvement, optimization

T. Najeh (✉) · A.J. Telmoudi · L. Nabli
Research Laboratory of Automatic Signal and Image Processing,
School of Engineers of Monastir, University of Monastir, Monastir, Tunisia
e-mail: najehtawfik@gmail.com

A.J. Telmoudi
e-mail: achraf_telmoudi@yahoo.fr

L. Nabli
e-mail: lotfinabli@yahoo.fr

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_9

and productivity. Therefore it is necessary to detect and isolate any failure to avoid damage that may be harmful in an environment where performance is paramount. As a result has appeared the field of fault monitoring. At the start monitoring focuses on areas that have a high level of risk as well as the nuclear arms industry sectors. But nowadays it is essential to adopt a performance monitoring module. To ensure the correct operating mode, the process of control and supervision required to accommodate continuous information of their instantaneous state. The accuracy of measurements is an important factor in the reliability of the control and monitoring system performance.

Monitoring methods will be compartmentalized into two main families: methods without mathematical models and those with. To the first family the quality of information can be achieved by improving the accuracy of measuring equipment and increasing the number of sensors. Because of cost and technical reasons, the choice of this solution, where several sensors are used to measure the same variable is limited to installations that have high technological risks. The second family is based on the redundancy of information and can be exploited to verify the accuracy of measurements. The advantage of these methods is the efficiency of detection and fault isolation; contrary to the analytical methods the cost of hardware installation should decrease.

The use of analytical redundancy techniques is based on finding the relationships shown in the measurements of variables to reach a mathematical model. It seems more and more difficult comes to large systems, with performance is less satisfactory. In contrast, a method based on redundancy as PCA allows exploiting the linear or non-linear relationships between those variables. Therefore all correlations are taken into account without an explicit form of the model inputs/outputs. PCA [1–4] is used in two steps, the first part provides the model obtained from the history of the system during normal operation, the second phase is the detection and isolation of faults by comparing the established behavior model and the observed. However, the detection phase has a delicate problem which has a significant impact on the precision of the model and its ability in failure classification. Consequently in the case of non-linear systems modeling PCA requires other tools help to set the optimal structure of the model. Artificial Intelligence approach (AI) is very effective to solve this problem. The use of AI on traditional linear PCA cannot solve non-linear problems.

In this paper, a new approach using GAs to estimate curve passing through the middle of probability distribution. Likewise, the use of this method is easily extended to the problem of detection and diagnosing data faults of nonlinear systems.

This work is organized as follows. The second section gives a description of principal curves. The third section introduces the concept of GAs and how we will apply this heuristic method to the problem of determining the principal curves. The new approach based on genetic and all steps are described in this section. In the last section the results obtained on numerical example is given.

2 Principal Curves

The first definition of principal curve is based on the self-consistency property of Hastie [2]. But this approach does not support closed curves and curves with intersections. A different method based on a model of the principal semi-parametric curve was proposed by Tibshirani [5]. But lack of flexibility it has the same weaknesses of the theory of Hastie.

Kègal [6] introduced anther definition based on the polygonal lines to find a principal curves. All the following approaches started research for a straight line, which by default is the first principal component [7, 8]. Another kind of approaches defined the principal curve with another way. Rather than starting with a line that represents the entire cloud of points, these approaches consider just a set of points. This principle was introduced by Delicado [3] for the construction of principal curves.

In this section we will analyze the self-consistency property of principal components. Let the data $X \in \Re$ generated, if $f(\lambda) = (f_1(\lambda), \ldots, f_d(\lambda))$, $\lambda \in \Re$ is the curve parameterized with $\lambda \in \Re$, then for any $X \in \Re^d$ we have $\lambda_f(X)$ the projection index and $f(\lambda)$ is a principal curve. Mathematically the projection index is defined by:

$$\lambda_f(\mathrm{X}) = \sup\left\{\lambda : \|\mathrm{X} - f(\lambda)\| = \inf_{\tau}\|\mathrm{X} - f(\tau)\|\right\} \tag{1}$$

For $X \in \Re^d$, the projection index $\lambda_f(X)$ is the largest value giving the minimum of $\|\mathrm{X} - f(\lambda)\|$.

3 Principal Curves Using Genetic Algorithms

The problem of determining the principal curves is a non-convex problem that has several possible solutions [9, 10]. To solve such problems, the classical approaches have multiple limitations. Due to the inadequate initialization of the algorithm or the predefined strategy of constructing local models, the obtained curves don't provide an optimal solution that is able to present, sufficiently, the complexity of the data cloud. On the other hand and despite the large computational cost, these methods do not allow a significant improvement for the construction of principal curves.

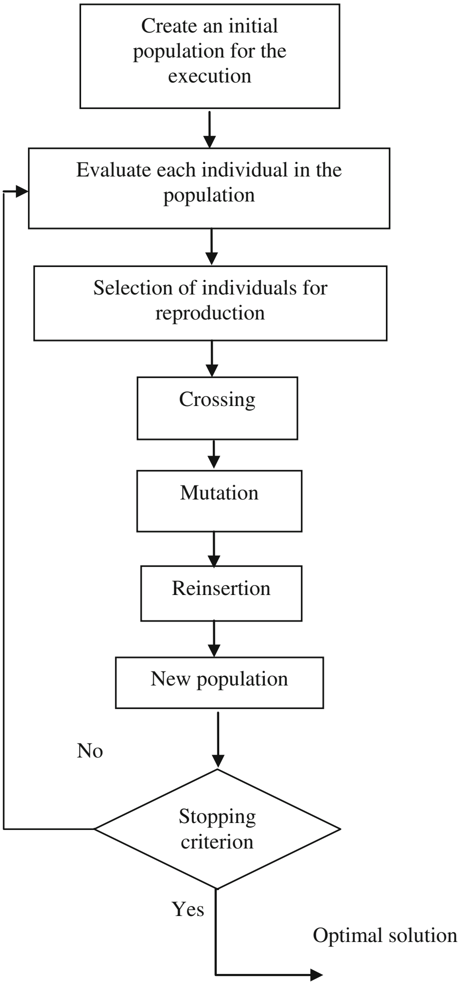
Using the genetic algorithms technique for the optimization of solutions of non-convex problem has attracted growing interest in many research works [11, 12]. The novelty of this technique is the assumptions commonly used with conventional methods to ensure convergence of the solution [13]. In the presence of multiple local optima, the convergence of GAs provides the desired solution to the global optimum of the problem [14]. The application of these tools in the case of the principal curve calculation is very interesting due to the non-convex nature of the problem.

3.1 Genitc Algorirthms

GAs are applied to a wide variety of problems. Simplicity and efficiency are the two advantages of this approach [12]. After having fixed the expression of the objective function to be optimized, probabilistic steps are involved to create an initial population of individuals [13]. Optimization steps with GAs are as follows (Fig. 1):

a. *Initialization*
   It is usually random and it is often advantageous to include the maximum knowledge about the problem [12].

Fig. 1 Chart of the simple genetic algorithm


Create an initial population for the execution
Evaluate each individual in the population
Selection of individuals for reproduction
Crossing
Mutation
Reinsertion
New population
No
Stopping criterion
Yes
Optimal solution

b. *Evaluation*

This step is to compute the quality of individuals by the allocation a positive value called "ability or fitness" to each one. The highest is assigned to the individual that minimizes (or maximizes) the objective function [14].

The fitness of an individual is calculated as follows:

$$\text{Fitness}(Pos) = 2 - P_s + \frac{2(P_s - 1)(Pos - 1)}{Nind - 1} \tag{2}$$

The evaluation is characterized by a parameter called selection pressure (P_s). This method allows P_s values in the range of [1, 2].

c. *The selection*

This step selects a definite number of individuals of the current population [13]. The selection is probabilistic; it is based on the ability of individuals a way that the best ones have a chance of being selected more than once. In this step is assigned to each individual probability P_i which is proportional to its fitness and defined by:

$$P_i = \frac{F_i}{\sum_{j=1}^{M} F_j} \tag{3}$$

With F_i the fitness and M the size of population.

d. *The crossover*

The genetic crossover operator creates new individuals. From two randomly selected parents, crossover produces two descendants [14]. This step affects only a limited number of individuals established by the crossover rate (Pc) number. Let $X = (x_i)_{1 \le i \le m}$ and $Y = (y_i)_{1 \le i \le m}$ be two individuals. These two parents will produce two offspring $X' = (x_i')_{1 \le i \le m}$ and $Y' = (y_i')_{1 \le i \le m}$ according to the equation:

$$\begin{cases} x_i' = x_i + s_i\, r_i\, a\, \frac{y_i - x_i}{\| Y - X \|} \\ y_i' = y_i + s_i\, r_i\, a\, \frac{x_i - y_i}{\| Y - X \|} \end{cases} \tag{4}$$

with: $a = 2^{-ku}$

k: mutation precision ($k \in \{4, 5, \ldots, 20\}$, $u \in [0, 1]$)

$$r_i = r \times domian$$

$$s_i = \{-1, 1\}$$

e. *The mutation*

The mutation consists in providing a small disruption to a number (Pm) of individuals. The effect of this operator is to counteract the attraction exerted by the best individuals this allows us to explore other areas of the search space.

Let u_i and l_i be the respective lower and upper bounds for all individuals. Let $X = (x_i)_{1 \leq i \leq m}$ the individual to mutate that will give the new individual $X' = (x'_i)_{1 \leq i \leq m}$ according to:

$$x'_i = \begin{cases} x_i + (l_i - x_i)f(G) & \text{if } r_1 < 0.5 \\ x_i - (x_i - u_i)f(G) & \text{if } r_1 \geq 0.5 \\ x_i & \text{if } x'_i \notin [u_i, l_i] \end{cases} \tag{5}$$

with:

$$f(G) = \left[r_2 \left(1 - \frac{G}{G_{\max}} \right) \right]^{b_s} \tag{6}$$

r_1, r_2: uniform random number between 0 and 1
G: the current generation
$G_{\max}$: the maximum number of generations
$b_{s:}$ shape parameter

To ensure the diversity of the population by the mutation the parameter r_1 is halved in Eq. 5 (r_1 > 0.5 and r_1 < 0.5).

This has been a brief overview of GAs. For further details on the processing power and the convergence properties of GAs, reference should be made to [15].

A. *finding principal curves*

The resolution of principal curve problem by GAs avoids all local optima and converges to the global optimum of the problem. The proposed approach considers the principal curves as an ensemble of connected lines segments. In each step new segment is inserted to form polygonal lines.

The use of GAs in order to find the principal curves requires the development of an objective function. This function must take into account the quadratic sum of the distances d_k.

$$\Phi_k = \sum_{i}^{k} d(x_i, s)^2 \tag{7}$$

where: X_n Two-dimensional random vector $x_i \in \Re^2$, i = 1,… k is a local neighborhood for a fix point "*a*"

s = [a, b]: a segment pass through the data set $x_{i.}$

d_j: the distance between the considered point and it's orthogonal projection given by Eq. (7).

The objective is used to find the segment that minimizes the total squared distance of all neighbor's points. We can project orthogonally every data point in the neighborhood cloud onto the segment s_k.

The new genetic curve algorithm is constructed following the strategy outlined as follows:

Algorithm:

```
 1: Start with a random point $x_0$ from $X_n$ .
 2: $L_n$ = the local neighbourhood of "$n$" points around $x_0$.
 3: Repeat until number generation = $Gmax$.
 4:   Generate at random an initial population of $K$ segments.
 5:       for every segment $s_k$ do
 6:           compute the Euclidean distance:
              $\phi_k = \sum_i^n d(x_i, s)^2$    for each segment $s_k$
 7:       end for
 8: Apply selection (Equation 3 )
 9: Apply crossover and mutation (Equations 4,5 ).
10: Save the better line segment and return to step 3.
11: Delete used local neighbourhood points from the original
    distribution.
12: The end point of previous segment is served as a starting
    point.
13: Connect with previous segment (Except first one).
14: Return 02 and repeat until all data points are achieved.
```

4 Experimental Results

This section describes how to use GAs that implements the proposed nonlinear PCA method for fault detection. An example is used to prove the performance of the proposed approach to find the principal curve for process monitoring.

The proposed algorithm has been tested with synthetic datasets. We conducted experiments on several artificial data set and with 2-d data space. Data points are disturbed along with Gaussian noise independently imposed on different dimensions of the given curve.

We start first with several typical synthetic datasets to test the aptitude of our algorithm for computation of principal curves, and then the monitoring problem of the Continuously Stirred Tank Reactor (CSTR) is investigated.

4.1 Synthetic Datasets

We generate different shaped curves, such arc-shaped and circle. Contaminated with small noise (0.03). The curve obtained by applying the algorithm is shown in Figs. 2 and 3. One can see that the curve is reconstructed quite well.

4.2 Nonlinear Monitoring of CSTR Benchmark

One of the most commonly used chemical reactors in the industry is the Continuously Stirred Tank Reactor (CSTR). [14].

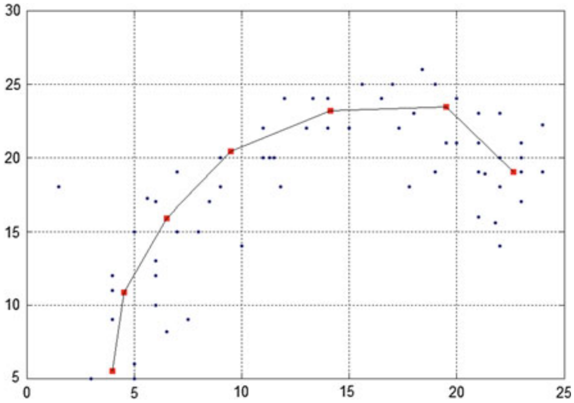


Fig. 2 Principal *curve* obtained for synthetic dataset of an *arc*

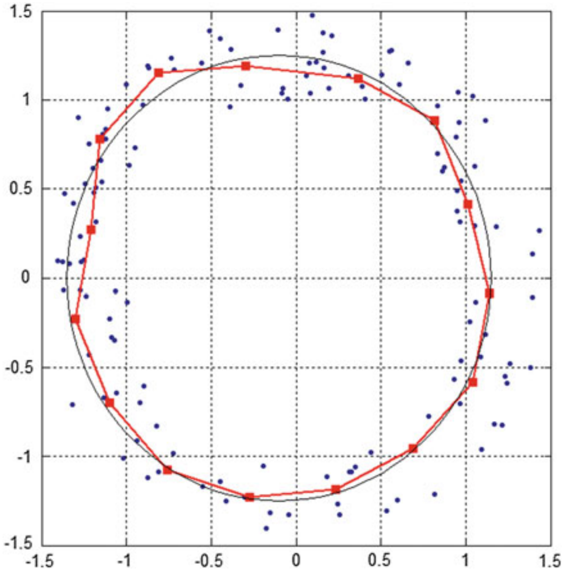


Fig. 3 Principal *curve* obtained for synthetic dataset of *circle*

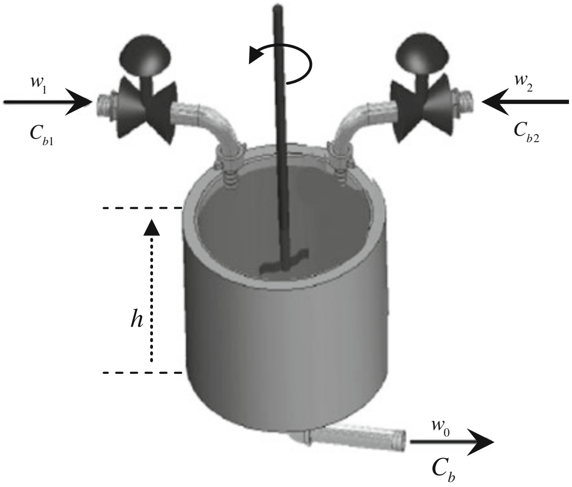


Fig. 4 Diagram of chemical reactor CSTR

Two chemical ingredients b1 and b2 come in the reactor with different concentrations and feed rates, respectively Cb1, Cb2 and w1, w2. This process produces the final product with the concentration Cb, feed rate w0 and the height h in the reactor. A diagram of the reactor is given in Fig. 4.

In this section we start by finding the principal curve of two process variables, $w_0 = x0$ and $h = x1$. second we propose it is use for the fault detection. The same Eq. 5 is taken as the objective function to be optimized. The initial population comprises 80 individuals randomly performed by the GP.

For linear process involving the linear approach of the PCA they use indices such as detection statistics of Hotling and SPE [16, 17]. Only in the case of non-linear PCA, application of these indices is not very suitable [18–20]. To overcome this difficulty, a new index detection I_f is proposed. The idea consists on constructing the principal curve of safe operating mode, then check index from the Euclidean distance between the estimated curve and data from the system at the present time.

For the previous example we considered two variables represented by a set of data points $x_n = (x_0, x_1)$. For each data point X_i let $p(x_i)$ being its projection point on the principal curve f. The Euclidean squared distance between x_i and $p(x_i)$ is calculated for the all data set. Then the deviation between estimating principal curve and the data set can be defined as:

$$I_f = d(x_i,f)^2 \tag{9}$$

Usually, the process is considered abnormal operating if:

$$I_f > \delta^2 \tag{10}$$

With δ is the threshold of detection.

In this study the principal curve was trained with a data set of 200 samples. The detection was performed with data containing one fault at a time. The test was designed for safe and failed operating mode.

To identify a change of the system's operating mode by the proposed method, we try to get the principal curve noted C_0 corresponding to normal operating mode on the absence of defects. This curve (Fig. 5) is obtained by the calculating algorithm of the principal curve of two variables (x_0, x_1).

From the constructed curve and the cloud of points, we can construct an indicator of change I_f through the Eq. (9). The process is simulated for 400 samples in the following manner; the first 250 samples correspond to the mode M_0 (Fig. 6). The second 150 ones correspond to mode M_I.

The evolution of I_f is shown in Fig. 3. In the interval [0, 250], the index I_f is below the threshold corresponding to the M_0 mode and above its threshold in the interval [251,400] corresponding to failed operating mode M_1.

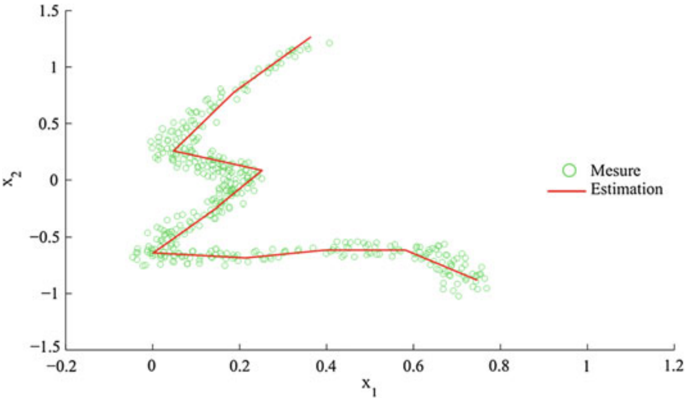


Fig. 5 Principal curve

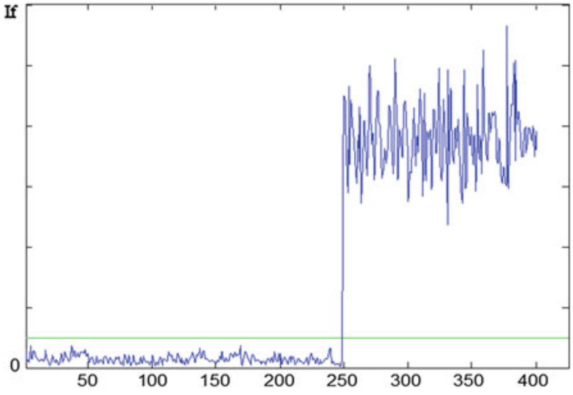


Fig. 6 Variation of indicator value I_f

5 Conclusion

In this paper a new approach was used to perform the principal curves based on genetic programming. The algorithm had been applied on some synthetic datasets and to the problem of monitoring on the Continuously Stirred Tank Reactor. The simulation improves the application of the proposed approach with real process. This guides us to try to apply this method with principal surfaces (higher dimension).

References

1. Nabli L (2010) Contribution à la conduite des systèmes de production par l'utilisation des techniques de l'intelligence artificielle. Habilitation universitaire, Université de Monastir, ENIM
2. Hastie T, Stuetzel W (1989) Principal curves. J Am Stat Assoc 84:502–512
3. Delicado P, Huerta M (2003) Principal curves of oriented points: theoretical and computational improvements. Comput Stat 18:293–315
4. Kègl B, Krzyzak A (2002) Piecewise linear skeletonization using principal curves. IEEE Trans Pattern Anal Mach Intell 24:59–74
5. Tibshirani R (1992) Principal curves revisited. Stat Comput 2:183–190
6. Kègl B, Krzyzak A, Linder T (2000) Learning and design of principal curves. IEEE Trans Pattern Anal Mach Intell 22:281–297
7. Banfield JD, Raftery AE (1992) Ice floe identification in satellite images using mathematical morphology and clustering about principal curves. J Am Stat Assoc 87:7–16
8. Chang K, Ghosh J (2001) A unified model for probabilistic principal surfaces. IEEE Trans Pattern Anal Mach Intell 23(1)
9. Einbeck J, Tutz G, Evers L (2005) Local principal curves. Stat Comput 15:301–313
10. Verbeek JJ, Vlassis N, Krose B (2002) A k-segments algorithm for finding principal curves. Pattern Recogn Lett 23:1009–1017
11. Goldberg D, Korb B, Deb K (1989) Messy genetic algorithms: motivation, analysis, and first results. J Complex Syst 3:493–530
12. Bies R, Muldoon F, Pollock G, Manuck S, Smith M (2006) A genetic algorithm-based hybrid machine learning approach to model selection. J Pharmacokinet Pharmacodyn 33:196–221
13. Cha SH, Tappert C (2009) A genetic algorithm for constructing compact binary decision trees. J Pattern Recogn Res 4:1–13
14. Akbari Z (2011) A multilevel evolutionary algorithm for optimizing numerical functions. Int J Ind Eng Comput 2:419–430
15. Zhang J, Chung H, Lo W (2007) Clustering-based adaptive crossover and mutation probabilities for genetic algorithms. IEEE Trans Evol Comput 11:326–335
16. Najeh T, Nabli L (2012) Development of a structuring residuals method for diagnostic by PCA-based genetic algorithms. In: The proceeding of 2nd international conference on communications, computing and control applications (CCCA), pp. 1–6, Marseilles, 6–8 Dec 2012
17. Nabli L, Ouni K, Haykel HS (2008) Approche Multi agents pour la surveillance indirecte d'un système de production par l'analyse en composantes principales. la Conférence Internationale Francophone d'Automatique, CIFA, Roumanie
18. Nabli L, Toguyéni AKA, Craye E (2000) Méthode de surveillance indirecte d'un système de production par la logique floue. CIFA, Lille

19. Kresta JV, MacGregor JF, Marlin TE (1991) Multivariate statistical monitoring of process operating performance. Can J Chem Eng 69:34–47
20. Nabli L, Ouni K (2008) The supervision indirect of a system of production by the principal component analysis and average dynamic of the metrics. Int Rev Autom Control (IREACO) 1 (4):560–567
21. Delicado P (2001) Another look at principal curves and surface. J Multi-variate Anal 7:84–116

Vibration and Acoustics Emissions Analysis of Helicopter Gearbox, A Comprative Study

Faris Elasha and David Mba

Abstract This paper investigates the application of signal separation techniques in detection of bearing faults within the epicyclic module of a large helicopter (CS-29) main gearbox using vibration and Acoustic Emissions (AE). It compares their effectiveness for various operating conditions. Three signal processing techniques including an adaptive filter, spectral kurtosis and envelope analysis, were investigated. In addition, this research discusses the feasibility of using AE in helicopter gearbox monitoring.

Keywords Vibration · Acoustics emission · Helicopter gearbox

1 Introduction

Helicopter transmission integrity is critical for safe operation. Approximately 16 % of mechanical failures, resulting in the loss of helicopter operation, can be attributed to the main gearbox (MGB) [1]. In addition, 30 % of the total maintenance cost of helicopters can be attributed to the transmission system [1]. The need to employ advanced fault warning systems for such transmission systems cannot be understated [2, 3]. Health and Usage Monitoring Systems (HUMS) are commonly used for fault detection of helicopter transmissions in which detection is based on extraction of predefined features of the measured vibration such as FM4, NA4, etc. [2, 4, 5]. HUMS was developed in North Sea operations, motivated in part by the crash to a Boeing Vertol 234 in 1986 which was caused by disintegration of the forward main gearbox. After development in the 1990s, the UK's Civil Aviation Authority CAA

F. Elasha (✉)
Faculty of Engineering, Environment and Computing,
Coventry University, Coventry CV1 5FB, UK
e-mail: faris.elasha@coventry.ac.uk

D. Mba
School of Engineering, London South Bank University, London SE1 0AA, UK
e-mail: mbad@lsbu.ac.uk

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_10

mandated fitment of HUMS to certain helicopters. One article suggests that HUMS "successes" are found at a frequency of 22 per 100,000 flight hours [6]. A HUM system consists of two complimentary subsystems: health monitoring and usage monitoring. Health monitoring is a process of diagnosing incipient damage or degradation that could ultimately lead to a system failure. Usage monitoring is a process by which the remaining life of different gearbox components and auxiliary systems is determined by assessing operation hours, current components condition and load history [7, 8]. Several vibration signature analysis methods are developed and applied in the commercial HUMS to detect faults in bearings, gears and shafts. Condition Indicators (CI) refer to the vibration characteristics extracted from these signatures and are used to reflect the health of the component [9]. Numerous condition indicators are calculated from vibration data to characterize component health and these indicators are often determined based on statistical measurement of the energy of the vibration signal, such as rms, kurtosis and crest factors.

The majority of helicopters utilises epicyclic reduction modules gears as transmission systems due to their high transmission ratio, higher torque to weight ratio and high efficiency [10]. As such this type of gearbox is widely used in many industries such as aerospace, wind turbines, mining and heavy trucks [11–15]. Different planetary gearbox configurations and designs allow for a range of gear ratios, torque transmission and shaft rotational characteristics. The planetary gearbox generally operates under severe conditions, thus the gearbox components are subject to different kinds of fault conditions such as gear pitting, cracks, etc. [16–19]. Recent investigations on applications of planetary gearboxes have shown that failures initiate at a number of specific bearing locations, which then progress into the gear teeth. In addition bearing debris and the resultant excess clearances cause gear surface wear and misalignment [19]. More recently the accident to the helicopter registred G-REDL [20], resulting in the loss of 16 lives, was caused by the degradation of a planet gear bearing interestingly the HUM system condition indicators showed no failure evidence before this accident.

2 Gear and Bearing Diagnostics

The vibration signals associated with bearing defects have been extensively studied and robust detection algorithms are now available as off-the-shelf solutions. Conversely the dynamics associated with bearing diagnostics within gearboxes reduce the effectiveness of traditional techniques. Therefore, it is important to understand the nature of the faulty bearing signal.

For rolling element bearings, a fault will cause shocks which in turn excite higher resonance frequencies which will be amplitude modulated depending on two factors, the transmission path and loading condition [21]. Therefore the vibration signal is typically demodulated to extract the frequency of these impulses. Equations for calculation of bearing faults frequencies have been reported widely in the literature [22–24]. These equations assume no slip, however, in operation there is some degree of slip and this why the bearing faults frequencies vary by 1–2 % of the calculated value. It is this slip that facilitates the separation of the gear and bearing vibration components [17], the latter known as the non- deterministic component of the measured vibration. The deterministic part of the signal is usually related to gear and shaft speeds [25]. Such periodic events are related to kinematic forces induced by the rotating parts such as meshing forces, misalignment and eccentricity [26]. In some cases the deterministic part of the vibration signal cannot be identified due to speed variation, and therefore it essential to re-sample the signal to the angular domain in order to track speed variation [26, 27]. The deterministic part of the signal can be used for diagnostics of gear and shaft faults.

In relation to AE only relatively short time series signatures were processed [28]. In application to diagnosis of machine faults, simple AE parameters are typically employed, such as rms, kurtosis, AE counts [29] and demodulation [30]. More recently the use of Spectral Kurtosis and adaptive filters has been employed to facilitate the diagnosis of machine faults with AE [31–33].

3 Signal Processing Techniques

Bearing and gear fault identification involves the use of various signal processing algorithms to extract useful diagnostic information from measured vibration or AE signals. Traditionally, analysis has been grouped into three classes; time domain, frequency domain and time-frequency domain. The statistical analysis techniques are commonly applied for time domain signal analysis, in which descriptive statistics such as rms, skewness, and kurtosis are used to detect the faults [34, 35]. A fast Fourier transform (FFT) is commonly used to obtain the frequency spectra of the signals. The detection of faults in the frequency domain is based on identification of certain frequencies which are known to be typical symptoms associated with bearing or gear faults. The time-frequency domain methods are composed of the short-time Fourier transform (STFT) [36], Wigner-Ville [34], and wavelet analysis [37, 38]. The use of these detection techniques are feasible for applications where a single component is being monitored however for applications that include several components, such as gearboxes, it is essential to employ separation algorithms. The adaptive signal processing techniques used in this study is fully described by the authors [39, 40].

4 Experimental Setup

Experimental data was obtained from tests performed on CS-29 Category ‘A’ helicopter gearbox which was seeded with defects in one of the planetary gears bearing of the second epicyclic stage. The test rig was of back-to-back rig configured and powered by two motors simulating dual power input.

4.1 CS-29 'Category A' Helicopter Main Gearbox

The transmission system of a CS-29 'Category A' helicopter gearbox is connected to two shafts, one from each of the two free turbines engines, which drive the main and tail rotors through the MGB. The input speed to the MGB is typically in the order of 23,000 rpm which is reduced to the nominal main rotor speed of 265 rpm, see Fig. 1.

The main rotor gearbox consists of two sections, the main module, which reduces the input shaft speed from 23,000 rpm to around 2,400 rpm. This section includes two parallel gear stages. This combined drive provides power to the tail rotor drive shaft and the bevel gear. The bevel gear reduces the rotational speed of the input drive to 2,405 rpm and changes the direction of the transmission to drive the epicyclic reduction gearbox module. The second section is the epicyclic reduction gearbox module which is located on top of the main module. This reduces the rotational speed to 265 rpm which drives the main rotor. This module consists of two epicyclic gears stage, the first stage contains 8 planets gears and second stage with 9 planets gears, see Fig. 2. The details of the gears are summarised in Table 1.

The epicyclic module planet gears are designed as a complete gear and bearing assembly. The outer race of the bearing and the gear wheel are a single component, with the bearing rollers running directly on the inner circumference of the gear. Each planet gear is 'self-aligning' by the use of spherical inner and outer races and barrel shaped bearing rollers (see Fig. 2).

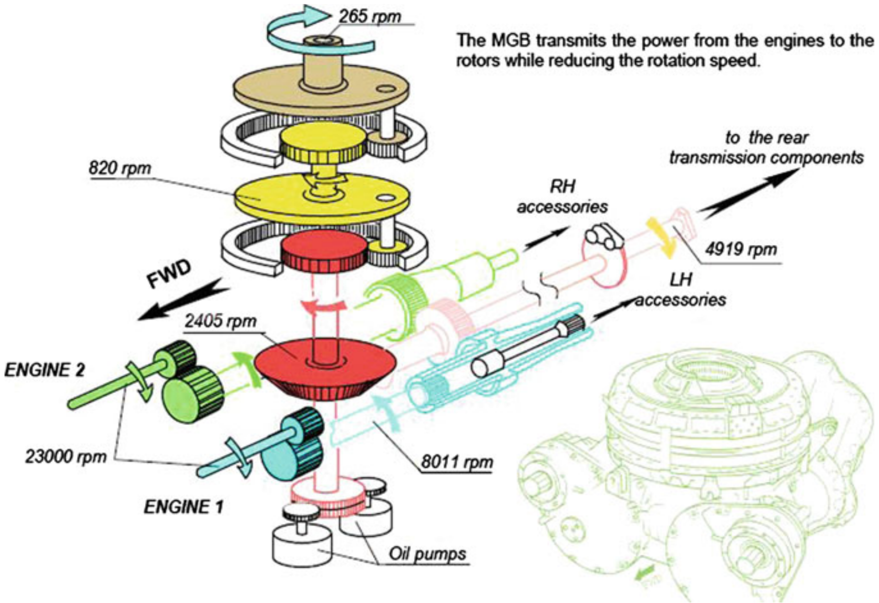

265 rpm
The MGB transmits the power from the engines to the rotors while reducing the rotation speed.
820 rpm
to the rear transmission components
RH accessories
4919 rpm
FWD
LH accessories
2405 rpm
ENGINE 2
23000 rpm
8011 rpm
ENGINE 1
Oil pumps

Fig. 1 Gearbox internal parts [20]

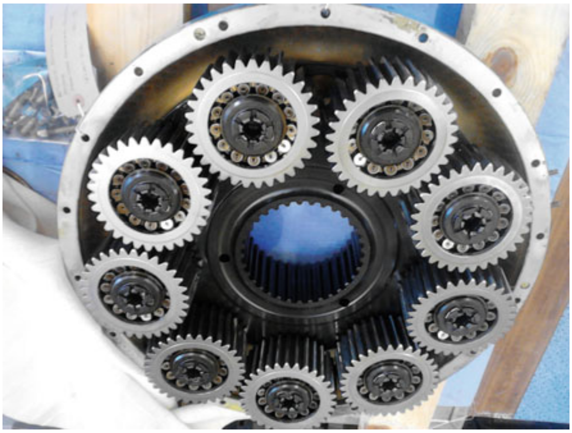


Fig. 2 Second stage epicyclic gears

Table 1 Number of teeth for the gearbox gears

First parallel stage	Pinion teeth	Wheel teeth	
	23	66	
Second parallel stage	Pinion teeth	Wheel teeth	
	35	57	
Bevel stage	Pinion teeth	Bevel teeth	
	22	45	
1st epicyclic stage	Sun gear	Planets gear—8 gears	Ring gear
	62	34	130
2nd epicyclic stage	Sun gear	Planets gear—9 gears	Ring gear
	68	31	130

4.2 Experimental Conditions and Setup

This investigation involved performing the tests for fault-free condition, minor bearing damage and major bearing damage. The bearing faults were seeded on one of the planet gears of the second epicyclic stage. Minor damage was simulated by machining a rectangular section of fixed depth and width across the bearing outer race (10 mm wide and 0.3 mm deep), see Fig. 3, and the major damage simulated as a combination of both a damaged inner race (natural spalling around half of the circumference) and an outer race (about 30 mm wide, 0.3 mm deep), see Fig. 4. Three load conditions were considered for the each fault condition, 110 % of maximum take-off power, 100 and 80 % of maximum continuous power; the power, speed and torque characteristics of these load conditions are summarised in Table 2.

Fig. 3 Slot across the bearing outer race

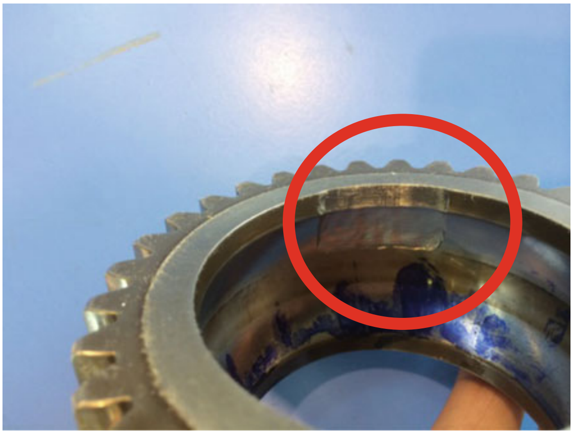


Fig. 4 Inner race natural spalling

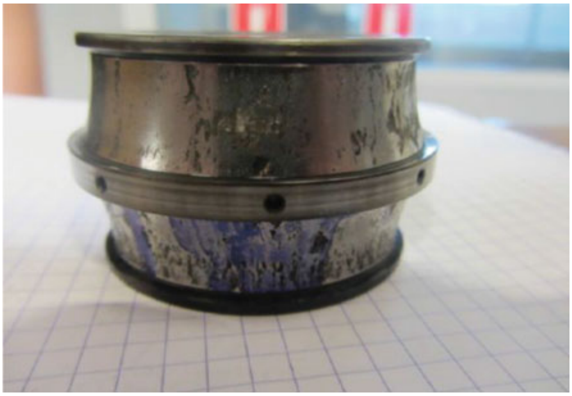


Table 2 Test load conditions characteristics

Load condition	Power (Kw)	Rotor speed (RPM)	Right input torque (Nm)	Left input torque (Nm)
100 % max continuous power	1300	265	272	272

4.3 Vibration Fault Frequencies

To aid diagnosis all characteristic vibration frequencies were determined, see Table 3. These included gears mesh frequencies of the different stages and the bearing defect frequencies for planet bearing.

Table 3 Gearbox characteristic frequencies

Frequency components	Frequency HZ
Gears meshes	
First parallel GMF Hz	8751.5
Second parallel GMF	4640.94697
Bevel stage GMF (Hz)	1791.24269
1st epicyclic stage GMF	1671
2nd epicyclic stage GMF	573
Faulty planet bearing	
Ball spin	45.31426
Outer race	96.69819
Inner race	143.9603
Cage	7.438322

4.4 Data Acquisition and Instrumentation

Vibration data was acquired with a triaxial accelerometer (type PCB Piezotronics 356A03) at a sampling frequency of the 51.2 kHz. The accelerometer had an operating frequency range of 2–8 kHz and was bonded to the case of the gearbox, see Fig. 5. The acquisition system employed was a National Instruments (NI) NI cDAQ-9188XT CompactDAQ Chassis. A 60 s sample was recorded for each fault case. The Y-axis of the tri-axial accelerometer arrangement was oriented parallel to the radial direction of gearbox, the X-axis to the tangential axis, and the Z-axis is the vertical axis parallel to the rotor axis, see Fig. 5.

In addition, Acoustic Emission data was collected using a PWAS sensor [41], 7 mm diameter and approximately 0.2 mm thick, bonded to the upper face of the planet carrier, see Fig. 6. The sensor was connected to a conditioning board attached to the planetary carrier and transmitted wirelessly using two coaxial copper coils and a new wireless transfer technique. The new wireless transfer technique utilise two single turn brass coils of approximately 400 mm diameter which were cut to size using water jets for accuracy. The stationary (upper) coil was suspended from two clamping rings which were attached to the top case of the gearbox with a spacer through the holes to retain location. The moving (lower) coil was attached to a circular mounting ring which was in turn mounted on top of the oil caps on the planet carrier, see Figs. 6 and 7. Electrical isolation of the coils from the mounts and surrounding metallic structure was achieved through the use of nylon washers and bushes. AE data was acquired at a sampling rate of 5 MHz using an NI PCI-6115 card connected to a BNC-2110 connector block.

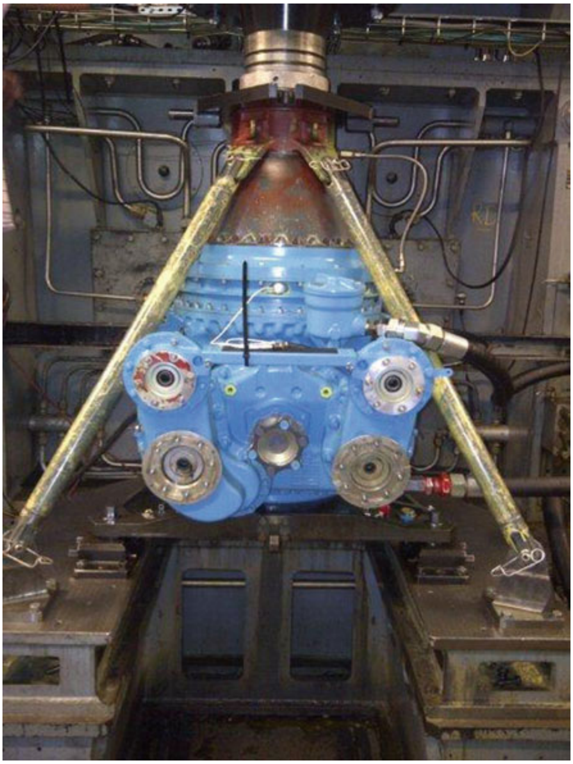


Fig. 5 MGB installed on the test bench

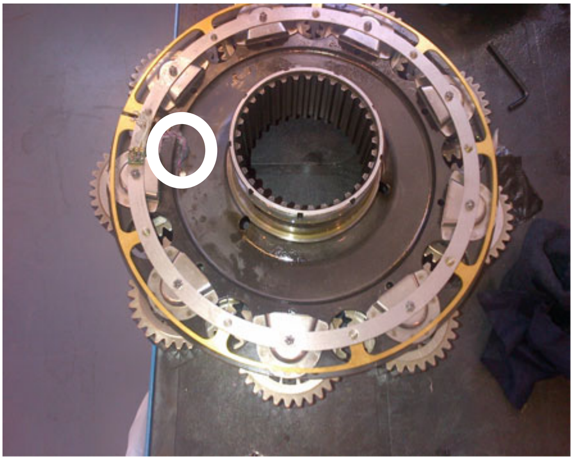


Fig. 6 Moving coil mounted on the planetary carrier (coil arrowed, sensor circled)

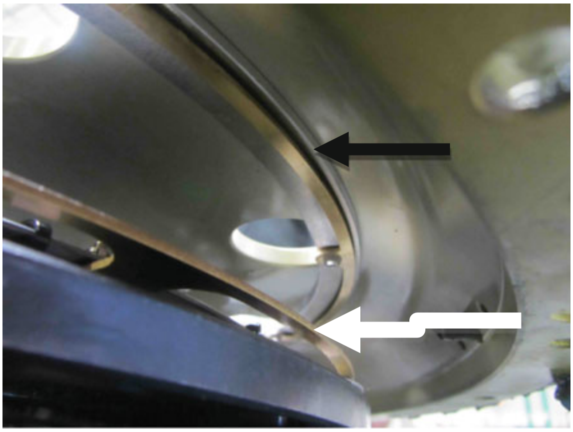


Fig. 7 Coils in position prior to assembly (static coil black arrow, moving coil white arrow)

5 Observations of Vibration Analysis

Spectral Kurtosis analysis was undertaken on the non-deterministic part of data sets collected from the gearbox for the different fault cases and this yielded the frequency bands and center frequencies which were then used to undertake envelope analysis. As discussed earlier the signal separation was undertaken with adaptive filter LMS algorithm. Observation from a typical Kurtogram used to estimate the associated filter characteristics for different defect conditions is shown in Table 4.

Observation from the spectra of the enveloped signal showed no presence of fault frequencies associated with the defective planetary bearing in the spectrum. However the minor fault condition was not identified. It is apparent that the signal separation had not completely removed the gear mesh and shaft frequencies, particularly the sun gears frequencies and its harmonics for first and second epicyclic stages (38.8 and 13.2 Hz respectively), which were detected by envelope analysis, see Fig. 8. Existence of these frequencies is due to fact that the vibration signal used in this analysis wasn't synchronised to any particular shaft.

Table 4 Filter characteristics estimated based on SK

Case	Center frequency Fc (Hz)	Band width Bw (Hz)	Kurtosis
Fault-free condition	5200	266	0.1
Minor damage condition	6000	266	0.11
Major damage condition	20266	2133	0.5

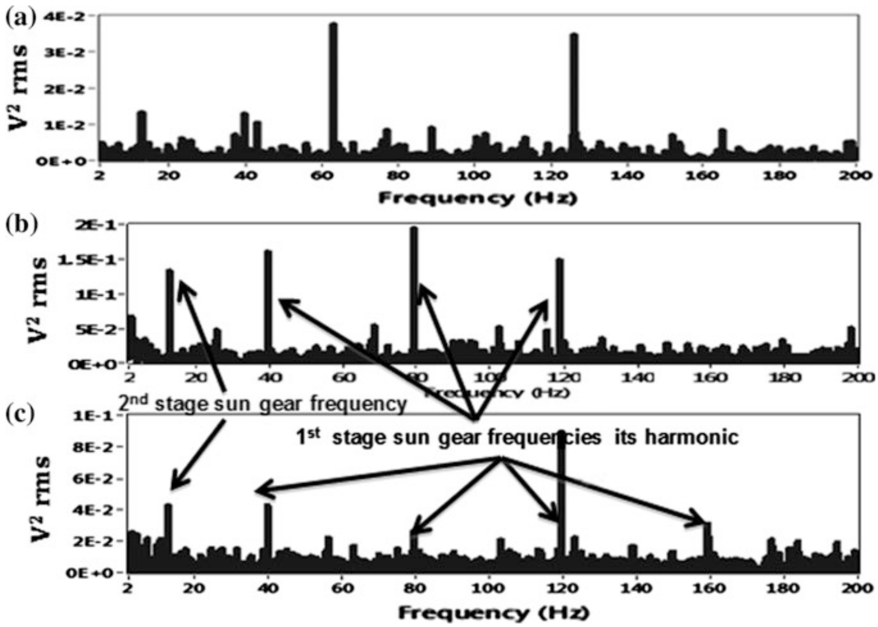


Fig. 8 Enveloped spectra of non-deterministic signal for (**a**) Fault-free (**b**) Major (**c**) Minor damage

6 Acoustic Emission Observations

The envelope analysis was undertaken using the central frequency F_c and bandwidth (Bw) estimated by SK analysis, see Table 5. Observations of Fig. 9 showed the presence of the bearing outer race defect frequency (96 Hz) and its harmonic (192 Hz) for both minor and major damages under different loading conditions.

Table 5 Filter characteristics estimated based on SK for AE signals

Case	Center frequency Fc (Hz)	Band width (Bw) (Hz)	Kurtosis
Fault-free	1093750	312500	12
Minor damage	234375	52083	9
Major damage condition	312500	208333	7.9

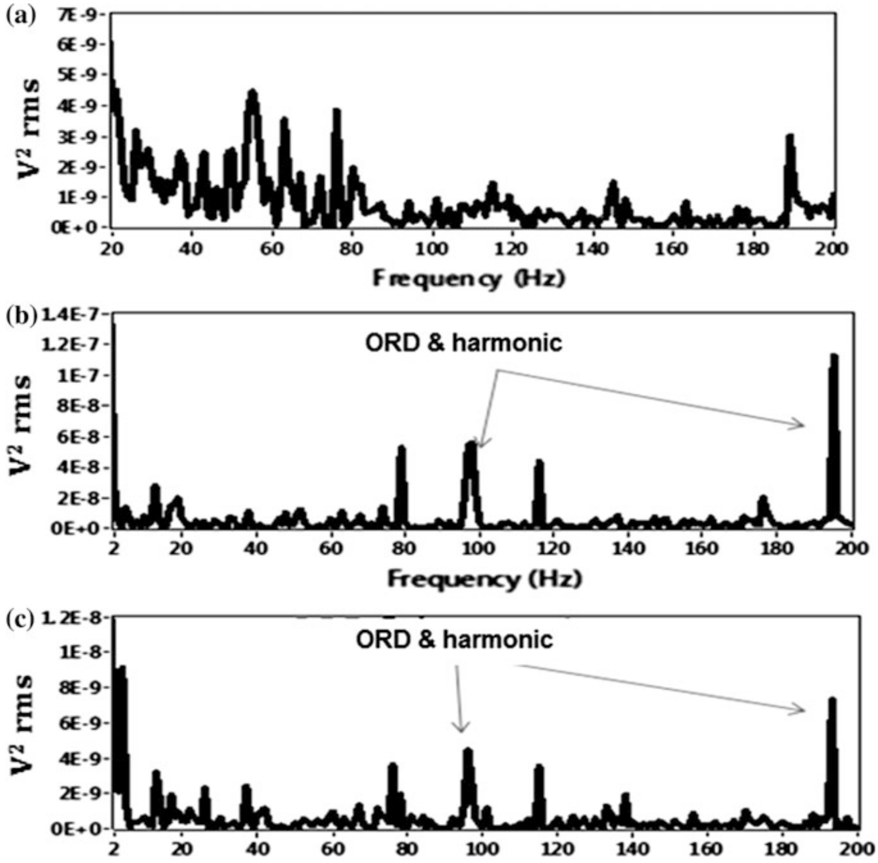


Fig. 9 Enveloped spectra of AE signal (**a**) Fault-free (**b**) Major (**c**) Minor bearing defects at 100 % maximum continuous power

7 Discussion

The techniques used in this paper are typically used for applications where strong background noise masks the defect signature of interest within the measured vibration signature. The AE signal is more susceptible to background noise and in this case, the arduous transmission path from the outer race through the rollers to the inner race and then the planet carrier makes the ability to identify outer race defects even more challenging. However the use of the wireless system incorporated into the main gearbox has contributed significantly to improving signal-to-noise ratio.

A comparison of the vibration and AE analysis showed AE analysis was able to identify the presence of the bearing outer race defect frequency (96 Hz) and its harmonic (192 Hz) for both minor and major damaged for all loading cases based

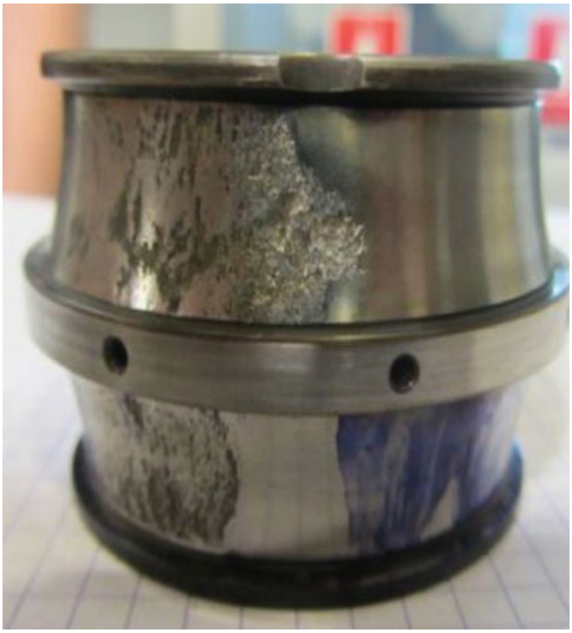


Fig. 10 Natural spall on bearing inner race

on observations on the enveloped spectra. However, for vibration analysis the outer race defect for minor damage case wasn't detected.

Interestingly the AE analysis was unable to identify the presence of a defective inner race under the 'major fault' condition however vibration analysis identified the presence of the cage frequency (7.5 Hz) for major fault condition. Under defective inner race conditions, as severe as that seen in this condition, see Fig. 10, it has been shown that such a fault condition manifests itself with increases in the bearing cage frequencies. The existence of large widespread spalls on the inner race leads to bearing excessive clearance which in turns causes an increase in the vibration amplitude of the fundamental train (cage) frequency.

8 Conclusion

In summary an investigation employing external vibration and internal AE measurement to identify the presence of a bearing defect in a CS-29 'Category A' helicopter main gearbox has been undertaken. A series of signal processing techniques were applied to extract the bearing fault signature, which included adaptive filter, Spectral Kurtosis, and envelope analysis. The combination of these techniques demonstrated the ability to identify the presence of the various defect sizes of bearing in comparison to a typical frequency spectrum. From the results presented it was clearly evident that the AE offered a much earlier indication of damage than vibration analysis.

Acknowledgement The Work described in this paper was conducted as part of EASA.2012. OP.13 VHM. The support of Airbus Helicopters in full-scale testing id gratefully acknowledged.

References

1. Chin H, Danai K, Lewicki DG (1993) Pattern classifier for health monitoring of helicopter gearboxes, No. NASA-E-7741, National Aeronautics and Space Administration (NASA) Cleveland OH Lewis Research Center, USA
2. Zakrajsek JJ (1994) A review of transmission diagnostics research at NASA Lewis Research Center, ARL-TR-599, NASA-TM-106746, E-9158, NAS 1.15:106746, NASA, USA
3. Chin H, Danai K, Lewicki DG (1993) Efficient fault diagnosis of helicopter gearboxes, (No. NASA-E-7975). National Aeronautics and Space Administration Cleveland OH Lewis Research Center
4. Decker HJ (2002) Crack detection for aerospace quality spur gears, NASA/TM—2002-211492, ARL-TR-2682. Glenn Research Center, NASA, USA
5. Zakrajsek JJ, Townsend DP, Decker HJ (1994) An analysis of gear fault detection methods as applied to pitting fatigue failure data. In: The systems engineering approach to mechanical failure prevention, vol 16, Apr 1993. Virginia Beach, Virginia, USA, NASA, USA, pp 199
6. Pipe K (2002) Measuring the performance of a HUM system-the features that count. In: Third international conference on health and usage monitoring-HUMS2003, pp 5
7. Samuel PD, Pines DJ (2005) A review of vibration-based techniques for helicopter transmission diagnostics. J Sound Vib 282(1–2):475–508
8. Decker HJ, Lewicki DG (2003) Spiral bevel pinion crack detection in a helicopter gearbox, NASA/TM—2003-212327—ARL–TR–2958. Glenn Research Center, NASA, USA
9. Dempsey PJ, Keller JA, Wade DR (2008) Signal detection theory applied to helicopter transmission diagnostic thresholds. In: Proceedings of the American helicopter society 65th annual forum on disc
10. Cotrell JR (2002) A preliminary evaluation of a multiple-generator drivetrain configuration for wind turbines. In: ASME 2002 wind energy symposium, American Society of Mechanical Engineers, pp. 345
11. Lynwander P (1983) Gear drive systems: design and application. CRC Press, Florida
12. Kahraman A (1994) Planetary gear train dynamics. J Mech Des 116(3):713–720
13. Huang C, Tsai M, Dorrell DG, Lin B (2008) Development of a magnetic planetary gearbox. IEEE Trans Magn 44(3):403–412
14. Radzevich SP (2012) Dudley's handbook of practical gear design and manufacture, 2nd edn. CRC press, USA. ISBN 9781439866016
15. Lu B, Li Y, Wu X, Yang Z (2009) A review of recent advances in wind turbine condition monitoring and fault diagnosis. In: Power electronics and machines in wind applications, 2009, PEMWA 2009. IEEE, p 1
16. McFadden PD (1987) A revised model for the extraction of periodic waveforms by time domain averaging. Mech Syst Signal Process 1(1):83–95
17. Randall RB (2004) Detection and diagnosis of incipient bearing failure in helicopter gearboxes. Eng Fail Anal 11(2):177–190
18. Wang W (2001) Early detection of gear tooth cracking using the resonance demodulation technique. Mech Syst Signal Process 15(5):887–903
19. Musial W, Butterfield S, McNiff B (2007) Improving wind turbine gearbox reliability. In: Proceedings of the European wind energy conference
20. Department for Transport (2011) Report on the accident to aerospatiale (Eurrocopter) AS332 L2 Super Puma, registration G-REDL 11 nm NE of Peterhead, Scotland, on 1 April 2009, 2/2011. Air Accident Investigation Branch, Aldershot

21. Randall RB, Antoni J (2011) Rolling element bearing diagnostics—a tutorial. Mech Syst Signal Process 25(2):485–520
22. Howard I (1994) A review of rolling element bearing vibration “detection, diagnosis and prognosis, DSTO-RR-0013, Department of defense
23. McFadden PD, Toozhy MM (2000) Application of synchronous averaging to vibration monitoring of rolling elements bearings. Mech Syst Signal Process 14(6):891–906
24. Khemili I, Chouchane M (2005) Detection of rolling element bearing defects by adaptive filtering. Eur J Mech A Solids 24(2):293–303
25. Sawalhi N, Randall RB, Forrester D (2014) Separation and enhancement of gear and bearing signals for the diagnosis of wind turbine transmission systems. Wind Energy 17(5):729–743
26. Antoni J (2005) Blind separation of vibration components: principles and demonstrations. Mech Syst Signal Process 19(6):1166–1180
27. Bonnardot F, El Badaoui M, Randall RB, Danière J, Guillet F (2005) Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation). Mech Syst Signal Process 19(4):766–785
28. Qu Y, Van Hecke B, He D, Yoon J, Bechhoefer E, Zhu J (2013) Gearbox fault diagnostics using AE sensors with low sampling rate. J. Acoustic Emission 31:67
29. Mba D, Rao RB (2006) Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines; bearings, pumps, gearboxes, engines and rotating structures
30. Holroyd T (2000) Acoustic emission as a basis for the condition monitoring of industrial machinery. In: Proceedings of the 18th Machinery vibration seminar, Canadian Machinery vibration association, pp 27
31. Ruiz-Cárcel C, Hernani-Ros E, Cao Y, Mba D (2014) Use of spectral kurtosis for improving signal to noise ratio of acoustic emission signal from defective bearings. J Fail Anal Prev 14 (3):363–371
32. Eftekharnejad B, Carrasco M, Charnley B, Mba D (2011) The application of spectral kurtosis on acoustic emission and vibrations from a defective bearing. Mech Syst Signal Process 25 (1):266–284
33. Kilundu B, Chiementin X, Duez J, Mba D (2011) Cyclostationarity of acoustic emissions (AE) for monitoring bearing defects. Mech Syst Signal Process 25(6):2061–2072
34. Sait A, Sharaf-Eldeen Y (2011) A review of gearbox condition monitoring based on vibration analysis techniques diagnostics and prognostics. In: Proulx T (ed) Rotating machinery, structural health monitoring, shock and vibration, vol 5. Springer, New York, pp 307–324. ISBN 978-1-4419-9427-1
35. Martin HR (1989) Statistical moment analysis as a means of surface damage detection. In: Proceeding of the 7th international model analysis conference, society of experimental mechanics, pp 1016–1021
36. Mehala N, Dahiya R (2008) A comparative study of FFT, STFT and wavelet techniques for induction machine fault diagnostic analysis. In: Proceedings of the 7th WSEAS international conference on computational intelligence, man-machine systems and cybernetics. Cairo, Egypt, World Scientific and Engineering Academy and Society, WSEAS; Stevens Point, Wisconsin, USA, pp 203
37. Wang WJ, McFadden PD (1996) Application of wavelets to gearbox vibration signals for fault detection. J Sound Vib 192(5):927–939
38. Wang WJ, McFadden PD (1993) Early detection of gear failure by vibration analysis i. calculation of the time-frequency distribution. Mech Syst Signal Process 7(3):193–203
39. Elasha F, Ruiz-Carcel C, Mba D, Chandra P (2014) A comparative study of the effectiveness of adaptive filter algorithms, spectral kurtosis and linear prediction in detection of a naturally degraded bearing in a gearbox. J Fail Anal Prev 14:1–14

40. Elasha F, Mba D, Ruiz-Carcel C (2015) Effectiveness of adaptive filter algorithms and spectral kurtosis in bearing faults detection in a gearbox. In: Sinha JK (ed) Springer International Publishing, pp 219–229. ISBN 978-3-319-09917-0
41. Yu L, Momeni S, Godinez V, Giurgiutiu V (2011) Adaptation of PWAS transducers to acoustic emission sensors. In: SPIE smart structures and materials nondestructive evaluation and health monitoring. International Society for Optics and Photonics, pp 798327

Test Rig Assessment of an On-Line Wear Sensor for Application in Wind Turbine Gearboxes

Vicente Macián, Bernardo Tormos, Santiago Ruiz, Guillermo Miró and Isaac Rodes

Abstract Wind energy is one of the most promising renewable energies, but it also presents some technical challenges, especially regarding to reliability, due to the cost of repair and maintenance actions. So, different solutions have been proposed from the point of view of on-line monitoring of gearbox condition by means of oil analysis. In this work, a complete process for the evaluation of a wind turbine gearbox on-line oil sensor was performed, based on a particle counting methodology. This work includes the design, selection and preparation of the samples studied and the test rig and all the experiments done for the assessment.

Keywords Oil analysis · Wind turbine gearbox · Particle counting

1 Introduction

In the last 20 years, the shortage of natural resources and the emerging social consciousness about pollution have made that renewable energy alternatives get an important and increasing role in energy production and specially wind energy

V. Macián · B. Tormos · S. Ruiz · G. Miró (✉)
CMT-Motores Térmicos, Universitat Politècnica de València,
Camí de Vera, s/n, València, Spain
e-mail: guimimez@mot.upv.es

V. Macián
e-mail: vmacian@mot.upv.es

B. Tormos
e-mail: betormos@mot.upv.es

S. Ruiz
e-mail: saruiz@mot.upv.es

I. Rodes
Tecnología de Materiales, Iberdrola Generación S.A,
Pol. Industrial "El Serrallo" s/n. Grao de Castellón, Castellón, Spain
e-mail: irodes@iberdrola.es

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_11

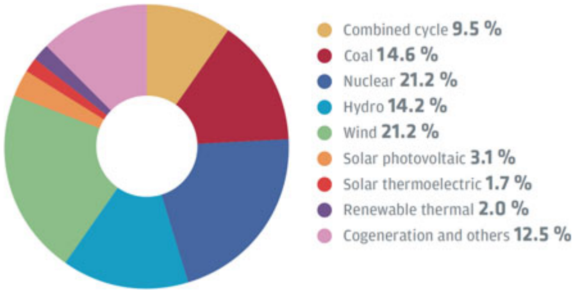


Fig. 1 Annual demand coverage for Spanish electrical system in 2013

among all of them. For instance, in the 2013 annual report of the Spanish electricity system [1], data regarding electricity coverage attending primary source are depicted (Fig. 1); showing clearly that wind energy is fighting to become the main energy source in Spain.

Therefore, leading companies in energy business have made economic investments in several different renewable energies (wind, solar, tidal, geothermal, hydro, etc.) that are increasingly present in our daily lives.

Among all of them, wind energy presents some advantageous characteristics. The power efficiency of a typical wind turbine is about 40 % (of the total kinetic energy of the wind), much more than solar energy (commercial panels available nowadays have approximately 20 % efficiency), furthermore wind turbines take up less space than the average power plant (a few square meters for the base), and this ground can be placed in remote locations, such as offshore, mountains and deserts. Additionally, combined with other alternative energy sources, wind can provide a reliable supply of electricity.

However, the implementation of renewable energies not only offers its profits to environmental and economic level, but also causes the appearance of new engineering challenges, including design, manufacturing and maintenance actions, that need sometimes a completely different approach.

A major issue with wind power systems is the relatively high cost of operation and maintenance (O&M). Usually, wind turbines are structures located in remote areas, presenting as a consequence a difficult access. Therefore, these factors increase the O&M cost for wind power systems. Also, poor reliability directly reduces availability of wind power due to the turbine downtime [2]. Regarding to maintenance actions, the main subassemblies in a wind turbine are the mechanical system, electronic control system, and electrical system, responsible to convert the kinetic energy in electricity. More common failures are originated in subsystems belonging to the mechanical system. This includes components such as: main shaft and bearings, gearbox, rotor brake, blades and generator.

Specially, one of the most important maintenance aspects in a wind turbine is the gearbox's condition, the element that converts low-speed, high-torque spinning from the blades into high-speed spinning for electrical energy conversion. Within the life cycle of the wind turbine, wind gusts lead to misalignment of the drive train and provoke gradual failure of the gear components. This failure is critical, as it creates a significant increase in the operating costs and downtime of a turbine, while greatly reducing its profitability and reliability. A gearbox replacement can cost up to 10 % of the original construction cost, enough to cut deep into the projected profits [3].

Existing gearboxes are a spinoff from marine technology used in shipbuilding, and an example is shown in Fig. 2.

Parallel to the design of these new engineering solutions, new challenges need to be observed. Field data [4] shows that the drive train and gearboxes of modern wind turbines (in the MW power production range) are the weakest part in the system, with great costs and mean time to repair (MTTR) associated. Thus, an important part of maintenance actions and efforts should focus in gearbox condition.

One challenge is the proper maintenance of these new systems and, in particular, the application of preventive and predictive maintenance principles for wear control in wind turbine gearbox by oil analysis. These systems present specific characteristics that make them really suitable for the application of on-line monitoring techniques. Particularly, these systems are usually installed in remote locations and need to be stopped for any maintenance inspection. These operations need to be realized by specially trained operator to climb to the nacelle with high security standards, and usually the mean time to repair comprises several hours. This situation implies a high economic cost for any off-line analysis, so different on-line monitoring techniques are being developed to help managers take better maintenance decisions.

As it was said, there are considerable challenges on the reliable operation of the system bearing and gear components [5]. Extremely miscellaneous conditions cause high contact stresses, generator faults and grid engagement cause impact loads and bearing skidding. On the other hand, ambient moisture causes corrosive environments and lubricant degradation. These conditions have resulted in issues of scuffing, micropitting, wear, pitting and surface cracking, as shown in Fig. 3.

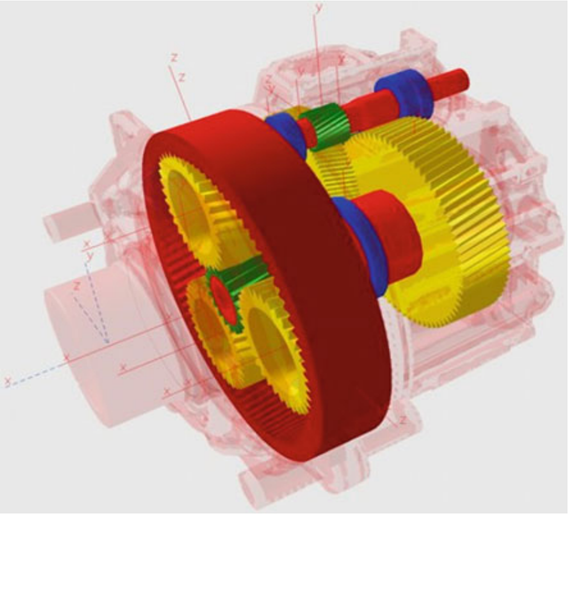


Fig. 2 Schematic wind turbine gearbox diagram. *Source* ZF Friedrichshafen AG

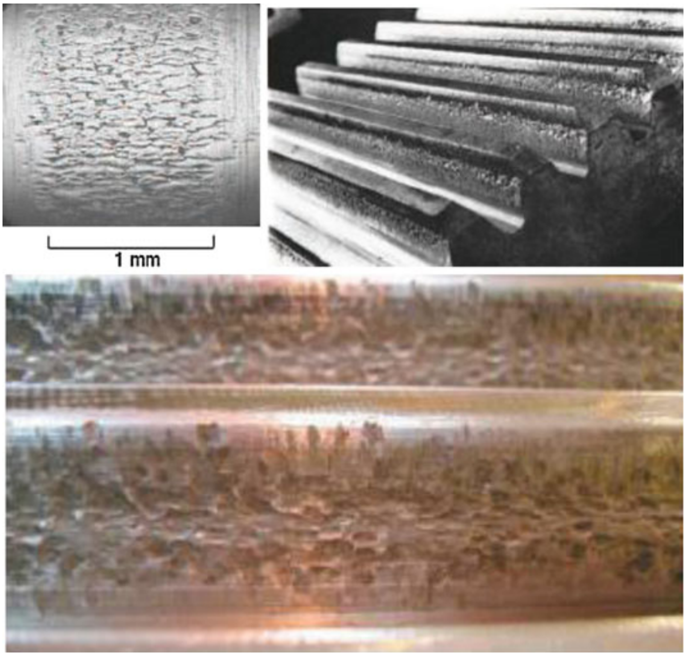


Fig. 3 Typical wear phenomena in wind turbine gearboxes. Micropitting (*upper left*), pitting (*upper right*), scuffing (*down*)

All phenomena mentioned above result in the appearance of wear debris in the oil. Depending on material and size, they can be classified within each type of wear, and also they can offer valuable information on the condition of the gearbox. In an experiment carried out with gears [6], the relationship between amount and size of particles in oil and gearbox condition was studied, with the results obtained in Fig. 4.

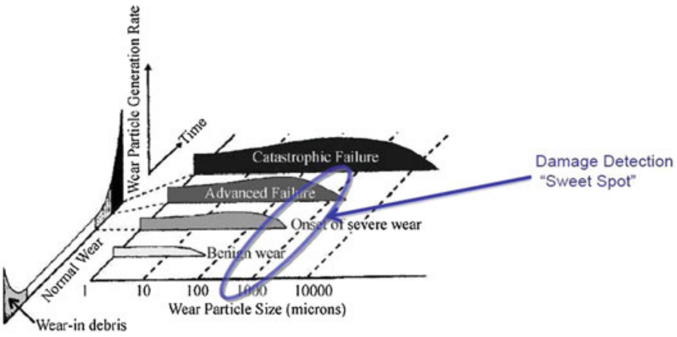


Fig. 4 Wear condition versus size and shape of particles in oil for gears [7]

The main target of this work has been the selection and subsequent test in laboratory of an on-line oil condition sensor to check if using this sensor an early stage detection of potential failures related with wear in a wind turbine gearbox can be achieved.

2 Design of Experiments

In order to develop this experiment, it was necessary to pay attention to three different aspects: selection and conditioning of the sensor, selection and manufacturing of the particles and design and development of the test rig.

2.1 Sensor

Particle counting devices represent one of the most important tests for used oil analysis, whether you use onsite particle counting or relying on a commercial lab performing off-line measurements. Thus some problems can be quickly and easily determined by monitoring the number and size-distribution of particles in an oil sample. Particle counting was introduced during the 1970s as a result of the pioneering work on hydraulics and fluid power conducted at Oklahoma State University, and then applied to a lot of different industries [8].

Many different particle counting principles are used nowadays, including optical and laser counting techniques. The principle used in this system needed to be robust and reliable, due to the difficulty of repairs once installed in the wind turbine. Thus, a magnetic detection system was selected. The sensor studied in this experiment was the MetalSCAN 3115L, designed by GasTOPS Ltd. (Canada), and shown in Fig. 5.

This sensor is an online particle counting sensor designed to detect and monitor metal particles in wind turbine gearbox oil, ferrous and nonferrous, generated by wear of the gearbox. The sensor generates an electric pulse for each particle larger than a minimum size, in order to detect the most interesting particle range. The detecting principle is based on the measurement of the magnetic field disturbance caused by the passage of a particle through the sensor. The particle is magnetized before entering the counter, then it is counted, and subsequently demagnetized before exiting the sensor. The waveforms that occur depend on the direction of motion of the particle and the type of particle [7], as shown in Fig. 6.

As said before, this sensor is connected directly to the lubrication system of the wind turbine. Table 1 summarizes the most important characteristics of the sensor.

The connection between the sensor and the PC was realized with a microcontroller with Ethernet Modbus TCP/IP protocol, which was needed to be configured in order to monitor and log the experiments.

Fig. 5 MetalSCAN 3115L

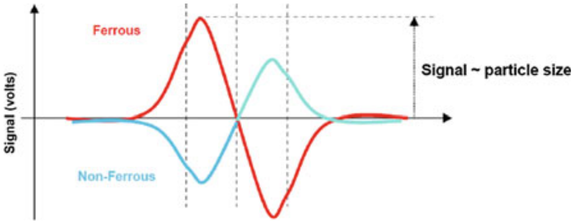


Fig. 6 Sensor signal for a particle, ferrous (*red*) and nonferrous (*blue*). Adapted from [7]

Table 1 Main characteristics of MetalSCAN 3115L

MetalSCAN 3115L	
Sensor bore	38 mm
Minimum size particle (spherical)	350 μm Fe/1000 μm Non-Fe
Minimum size particle (equivalent spherical diameter)	230 μm Fe/600 μm Non-Fe
Flow rate	38–1000 l/min

2.2 Particles in Oil

Once the sensor was selected, next step comprised the selection and preparation of particles and oil necessary for the experiment.

First of all, the oil selected was a typical commercial PAO-based (poly-alpha-olefins) ISO VG 320, whose main characteristics are described in Table 2.

Then, the selection of particles was carried out. According to American Gear Manufacturers Association (AGMA) [9], the main metallic elements present in oil are iron (Fe) and copper (Cu), as gears are mainly manufactured in iron and copper. According to this and taking into account the sensor limits, it was decided that two different types of particles will be studied. Initially, it was decided to get a particle size range higher than the sensor minimum, in order to assure the validity of measurements, thus two different materials were acquired (from Alfa Aesar):

- "*Iron powder, −20 mesh*": Iron particles with size less than 840 µm. A sample of these particles was studied in a laser granulometer in order to confirm size and verify the percentage of particles that could be used in the study, as shown in Fig. 7. After that, particles were sieved to a range greater than 300 µm.
- "*Copper shot, 0.6–0.8* mm": These particles present spherical to hemispherical form with different sizes, in this case, ranging between 600 and 800 µm, so no sieving was necessary. In Fig. 8 the particle size distribution of this sample is presented.

Table 2 Main properties of ISO VG 320 oil used in the experiment

Properties	Value
Kinematic viscosity @ 40°C [cSt]	320
Kinematic viscosity @ 100°C [cSt]	34.1
Density @ 15°C [kg/m^3]	853
Flashpoint, Cleveland open cup [°C]	260
Pour point [°C]	−54
Neutralization number [mgKOH/g]	0.6

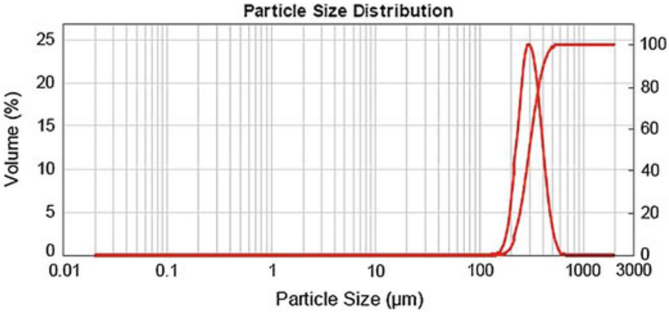


Fig. 7 Particle size distribution of "Iron powder, −20 mesh"

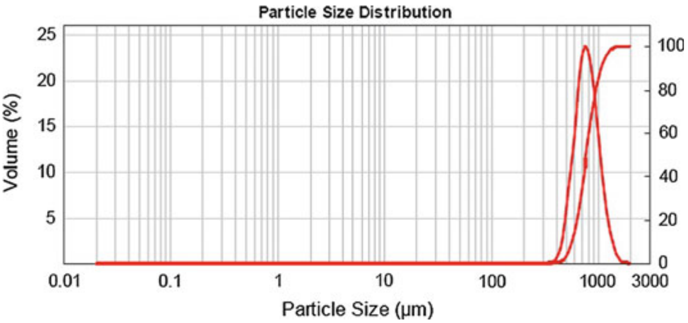


Fig. 8 Particle size distribution of "Copper shot, 0.6–0.8 mm"

Additionally, it was decided to transform a portion of these particles in a "flake" shape, since this is the usual shape for wear particles in a wind turbine gearbox.

In order to simulate this difference, a sample of these particles was milled. In Fig. 9 the difference between spherical particles and milled particles is shown.

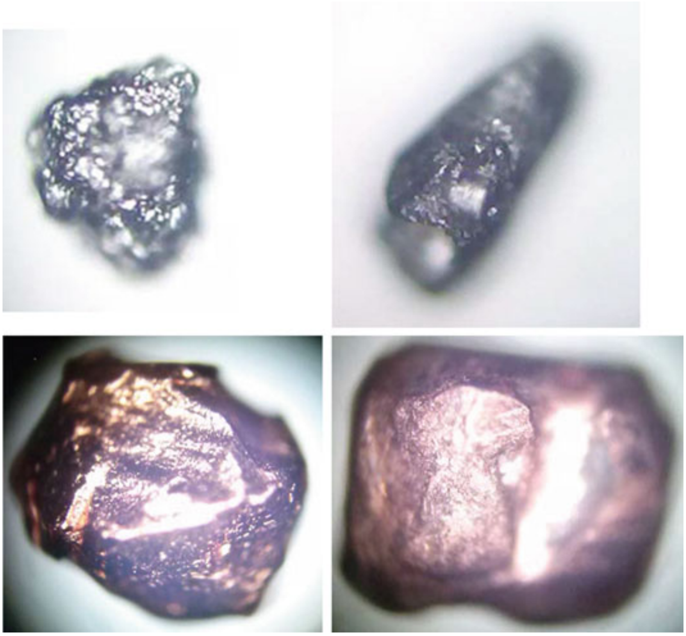


Fig. 9 Particle milling: original and milled iron particles (*up*) and original and milled copper particles (*down*)

2.3 Test Rig

For the purpose of this experiment, a test rig was designed especially for the sensor assessment. In Fig. 10 a diagram of the test rig is presented.

The test rig configuration was designed according to the following procedure: the oil is in a deposit, where it is stored. Before starting the cycle, the oil is filtered by a submerged filter to prevent undesirable particles flowing through the sensor. The oil flows up to the pump, and after that the particles are introduced, in order to protect the pump from wear induced by the particles. Once the flow of oil and particles pass through the sensor, there is a magnet, whose function is to collect the particles used in the experiment, so the oil gets back clean to the deposit. One of the main points addressed was the selection of the pump, since there was a minimum flux rate, and it was needed also some flexibility, since the viscosity of oil is around 200 cSt at the temperature of normal operation. In Fig. 11 can be observed the final test rig completely assembled.

One of the most important parts that were designed included the particle injection system. After the evaluation of different alternative systems, a system based on a tee coupler was implemented, where the particles would be dragged by the oil impulse by the pump, shown in Fig. 12.

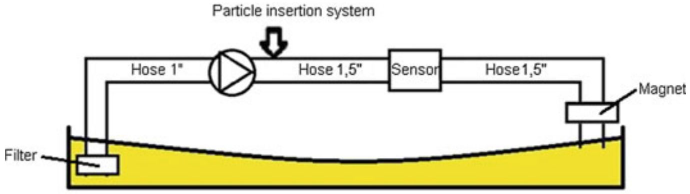
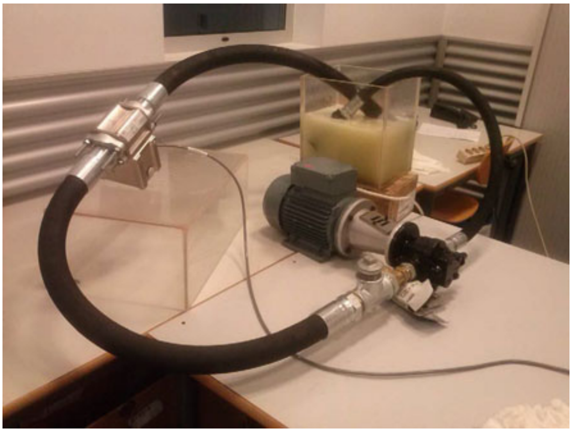


Fig. 10 Diagram of the test rig specifically designed for this experiment

Fig. 11 Mechanical assembly of the test rig



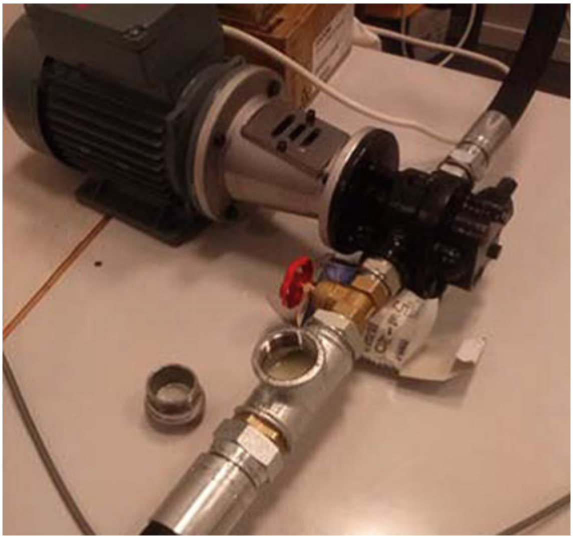


Fig. 12 Particle injection system

2.4 Experiments

Once the system was assembled and prepared, the sequence of experiments were decided. First, it was decided that the first experiment should be conducted with iron particles, after that with copper particles and finally using a mixture of them. The last option was included to increase the similarity of the test with real world conditions, since it would be the most common situation.

Thus, a blank test would be conducted, and later on, the sensor would be tested having a single particle. After that, a sensibility test would be carried out to find out the limitations of the sensor. Finally, it would be mandatory to realize some tests in real conditions to confirm the sensor capability. In Table 3 the different test performed are presented.

For the selection of incipient and severe wear, information was obtained from ANSI [9], and from similar tests [7], as shown in Fig. 13.

Table 3 Test sequence of the experiment

Experiment	Material	Particle rate
Blank	–	–
2	Fe	1
3	Cu	1
Sensibility test	Cu	1/min; 2/min; 3/min; 6/min; 12/min; 60/min
5	Fe+Cu	Incipient wear
6	Fe+Cu	Severe wear

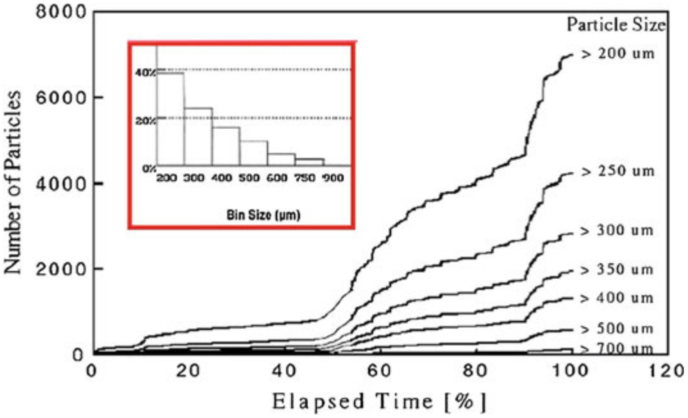


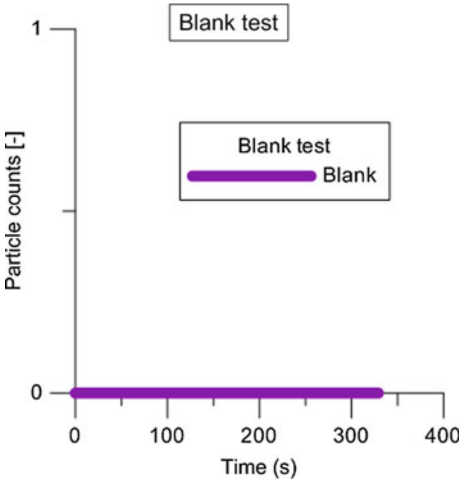
Fig. 13 Particle size distribution and rate of a typical gear failure. Adapted from [7]

With the information obtained, different masses of each element were considered for incipient and severe wear, according to the number of particles, oil flow rate, mean spherical diameter and density of each element.

3 Results and Discussion

First test performed was a blank test, in order to validate the normal operation of the sensor in the test rig and the connections realized. Results of this experiment are presented in Fig. 14.

Fig. 14 Blank test results



The result showed no counts, as expected. After that, a test using just one single particle for each metal considered was performed. The test was performed according to the following procedure: once the test rig was turned on, one particle was added through the particle injection system until the particle was detected. In Fig. 15 the results for the metals considered are presented.

Also in this case, the results showed expected trends. After the basic tests, the sensibility test was performed. As the test was being conducted, it was observed that there was some difficulty to introduce particles at a high speed rate, above 30 part/min. For that reason, the latter sensibility test was performed at the very end of the experiment, with the sensor extracted from the test rig, and the particles were passed through the sensor by free falling. The results obtained are shown in Fig. 16.

In Table 4 the rate comparison is found.

Main results from the sensibility test were that the sensor responded to a very wide range of functioning, and as the rate was increased, the sensor started to miss some counts. This response appeared as a consequence of signal counting algorithm design, since it was configured to detect faults in real environment of a wind turbine. If the experiment was performed with a great accumulation of wear particles in a short period of time, the sensor software configuration detects this phenomenon as inadequate environmental conditions and therefore rejects these measurements in order to avoid false-positive wear warnings.

With all the useful information obtained from the other test, a "real conditions" test was performed. According to the specifications given above, a quantified amount of both type of particles were introduced in the flow, one corresponding to incipient wear, and the second one simulating severe wear. Results are shown in Fig. 17.

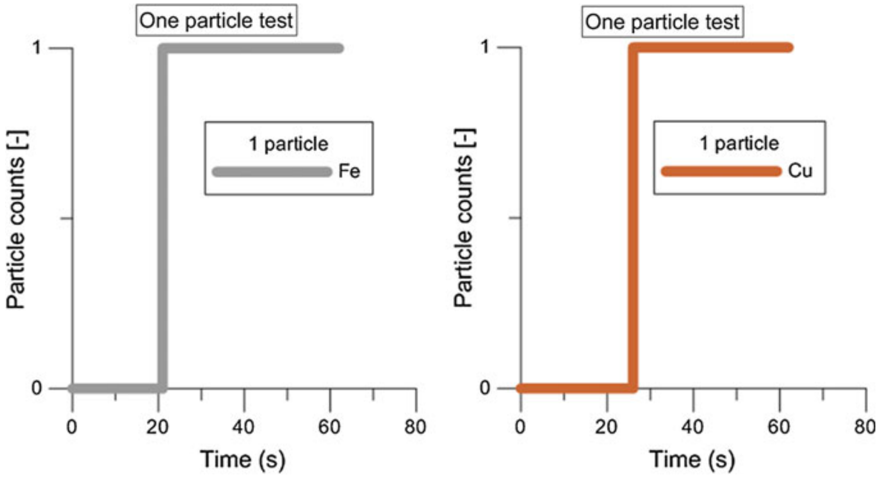


Fig. 15 One particle test results, for iron (*left*) and copper (*right*)

Fig. 16 Sensibility test results

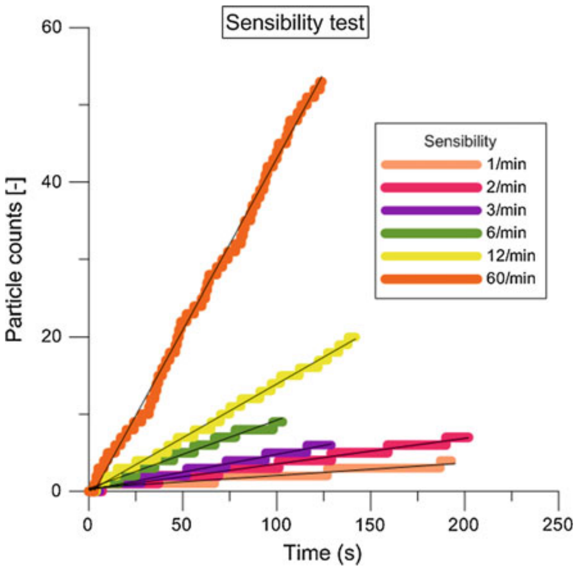
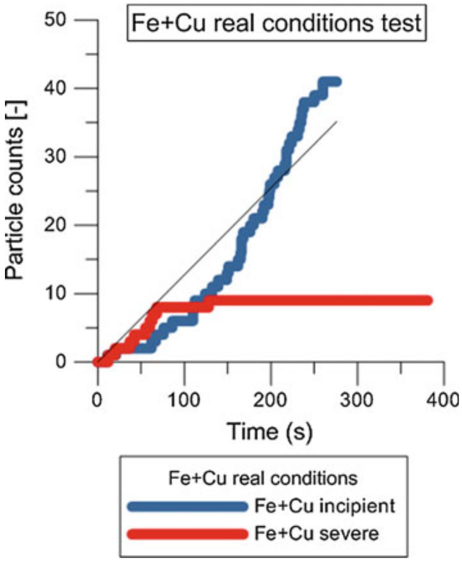


Table 4 Sensibility test rate comparison

Rate (per min)	Theoretical [counts/s]	Test [counts/s]
1	0.020	0.017
2	0.035	0.033
3	0.048	0.050
6	0.09	0.10
12	0.14	0.20
60	0.43	1.00

Fig. 17 Real conditions test results



These results confirmed the trends expected. In the particle injection system the particles were introduced manually into the flow, so at any time it was easy to introduce a large amount of particles in the system, and afterwards activating the false-positive algorithm of the sensor.

4 Conclusions and Future Works

The main conclusions of the experiment realized are detailed below:

- In this experiment a test rig was developed, to simulate the environment and characteristics of a lubricant circuit of a wind turbine gearbox, including oil selection and flow properties.
- A complete set of particles, both ferrous and non-ferrous, were selected and transformed in order to simulate typical wear particles.
- The sensor studied in this experiment showed good performance, detecting both types of particles and showing good sensibility test, for the usual range of particle rates.
- In simulated real conditions, introducing a particle amount corresponding to severe wear result a greater amount of particles than the amount corresponding to incipient wear, but the limitations of the particle injection system led to saturated measurements.

Considering these results, some future works have been proposed:

The first task to do would be the use of this test rig with real oil samples with signs of wear at different stages (from lowest to highest severity of wear) from real wind turbines on field. Thus, the sensor could be evaluated with real particles, in terms of number, shape and sizes.

Another future work, practically mandatory before a general installation in a fleet of wind turbines, would be to validate the behavior of the sensor studied under real working conditions, i.e. installed into the lubrication system of a little group of wind turbine gearboxes, preferably in those ones that have shown interesting wear trends in off-line analysis.

5 Acknowledgments

The development and success of this project could not be possible without the hard work and commitment of Eduardo Iniesta, and authors would like to thank very specially to him his help and efforts during the whole project.

Additionally, authors would like to thank Mr. Daniel Obedi. Also, we would like to extend this gratefulness to Vicente Amigó and Jenny Zambrano for their help in selection and transformation of particles, and to María Victoria Borrachero from ICITECH for the characterization of them.

Also, authors would like to recognize the work of Luis Navarro and Lorena Garzón in this project. Finally, authors would like to thank Iberdrola Generación S. A. for their support in the project.

References

1. REE (Red Eléctrica de España) (2013) The Spanish electricity system 2013. http://www.ree.es/es/node/5189. Accessed 25 Nov 2014 (online)
2. Lu B, Li Y, Wu X, 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 2009). IEEE, pp 1–7
3. Ragheb A, Ragheb M (2010) Wind turbine gearbox technologies. In: Proceedings of the 1st international nuclear and renewable energy conference (INREC10), pp 1–8
4. Sheng S, Veers PS (2011) Wind turbine drivetrain condition monitoring—an overview. National Renewable Energy Laboratory
5. Greco A, Sheng S, Keller J, Erdemir A (2013) Material wear and fatigue in wind turbine systems. Wear 302(1–2):1583–1591
6. Dempsey PJ, Lewicki DG, Decker HJ (2004) Investigation of gear and bearing fatigue damage using debris particle distributions. No. NASA-GRC-E-14297. National Aeronautics and Space Administration, Glenn Research Center, USA
7. Dupuis R (2010) Application of oil debris monitoring for wind turbine gearbox prognostics and health management. In: Proceedings of the annual conference of the prognostics and health management society, pp 10–16
8. Fitch EC (1988) Fluid contamination control, vol 4. FES Incorporated
9. American National Standards Institute (2004) Standard for design and specification of gearboxes for wind turbines (ANSI/AGMA/AWEA 6006-A03)

Residual Signal Techniques Used for Gear Fault Detection

Omar D. Mohammed and Matti Rantatalo

Abstract The role of vibration monitoring is to detect any impact on the vibration signal due to gear degradation and to give an early warning. Early detection allows a proper scheduled shutdown to prevent failure. Residual signal method can be applied to improve the extraction of the hidden fault impact. The current paper presents a comparative study of three different residual techniques. The paper concludes with a brief discussion on the used methods.

Keywords Gear fault detection · Gear dynamics · Residual signal method

1 Introduction

Gears are widely used in different applications for mechanical power transmission. Gear failure can occur due to an excessive applied load, insufficient lubrication, manufacturing errors or installation problems. In gear systems the vibration signal is dominated by the gear meshing vibration, which is accompanied by some amount of noise and probable geometric and assembly errors. Additional impacts will be present in the signal when a localized gear fault occurs. The additional impacts due to the existence of a fault are masked by the regular signal components. To improve the extraction of the hidden fault impact, residual signal method can be applied.

Thus, the idea of generating a residual signal is to remove the regular signal components in order to detect the fault more effectively. Different techniques have been developed in the past for generating the residual signal. The first technique was basically proposed by Stewart [1], who developed a number of fault detection

O.D. Mohammed (✉) · M. Rantatalo
Division of Operation and Maintenance Engineering,
Lulea University of Technology, Luleå, Sweden
e-mail: omar.mohammed@ltu.se

M. Rantatalo
e-mail: matti.rantatalo@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_12

indicators. Stewart's enhancement technique of obtaining a residual signal involves the removal of the gear mesh harmonics from the spectrum. Later on, Wang and Wong [2] developed a new filtering technique based on the autoregressive model (AR model). In this technique, the filtered signal which describes the healthy case was subtracted from the unfiltered signal to produce the AR model residual signal. The authors presented results showing that the AR model was more efficient and could detect a fault earlier than the traditional technique of Stewart. More recently, another method for removing the regular components was applied in Refs. [3–5], which involves the subtraction of the whole vibration time signal of the healthy case from that obtained with the existence of a fault. The rest of the signal was then the residual signal which contained information supposed to be only related to the gear fault. Finally, a residual signal technique based on the ensemble empirical mode decomposition (EEMD) method was proposed in Ref. [6]. Using this technique, the residual signal was obtained by removing some intrinsic mode functions (IMFs) which represent the meshing frequency harmonics and the other regular signal components.

In the current paper, three different residual techniques were applied to a vibration signal to compare their behaviour. The analysed vibration response was obtained by simulation using a gear dynamic model.

2 Gear Modelling

A program was developed using Matlab™ to investigate the time-varying gear mesh stiffness analytically. A crack case of a 1 mm crack depth has been modelled. Modelling of gear tooth crack is shown in Fig. 1. The main gear modelling parameters that were used for stiffness calculations were adopted from Refs. [5, 7, 8], and can be seen in Table 1.

A dynamic simulation of a 6 DOF model was performed based on the time-varying gear mesh stiffness value. Figure 2 shows the dynamic model which

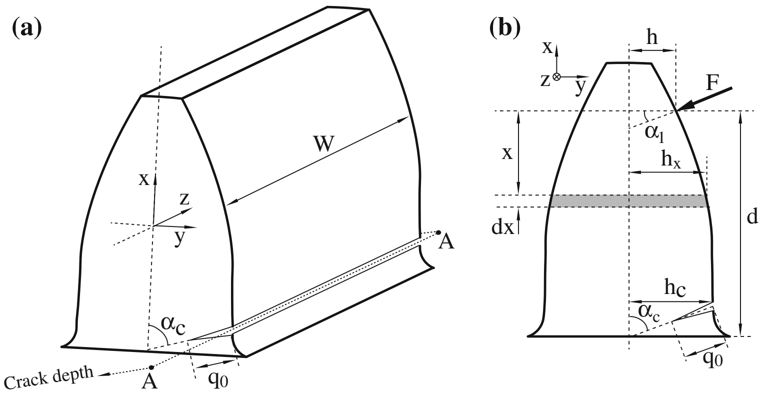


Fig. 1 Modelling of gear tooth crack. **a** modelling of cracked tooth, **b** tooth notation

Table 1 Parameters of gear-pinion set

Parameter	Gear	Pinion	Parameter	Gear	Pinion
Number of teeth	30	25	Mass (kg)	0.4439	0.3083
Module (mm)	2	2	Mass moment of inertia (kg.m^2)	2×10^{-4}	0.96×10^{-4}
Teeth width (mm)	20	20	Radial stiffness of the bearing in x,y direction (N/m)	6.56×10^8	6.56×10^8
Contact ratio	1.63	1.63	Radial damping of the bearing in x,y direction (N/m)	1.8×10^3	1.8×10^3
Rotational speed (rpm)	2000	2400	Coefficient of friction	0.06	0.06
Pressure angle (deg.)	20	20	Total damping between meshing teeth (N.s/m)	67	67
Young's modulus, E (N/mm^2)	2×10^5	2×10^5	Poisson's ratio	0.3	0.3

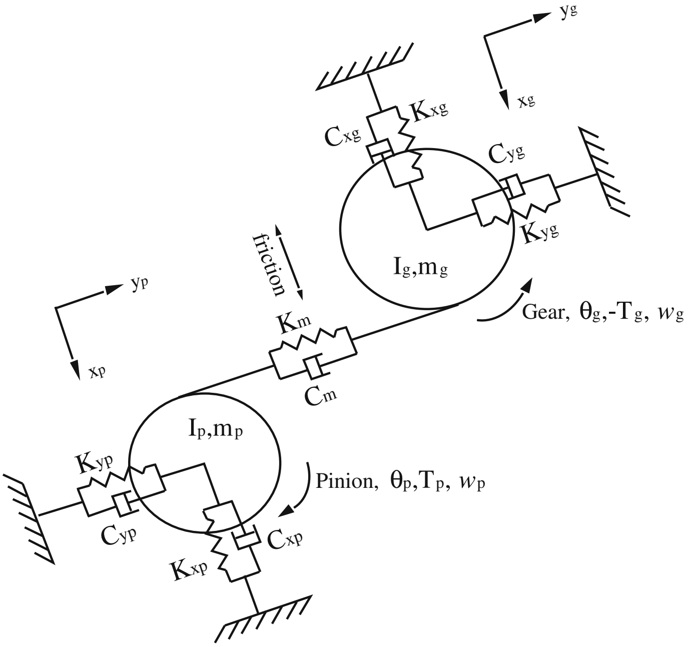


Fig. 2 Dynamic model of a one-stage gear system with 6 DOF

was used in the current research study and which was adopted in Refs. [5, 7, 9, 10]. A Matlab™ computer simulation using the ODE45 function was used for modelling the equations of motion. The dynamic simulation was performed for the healthy case, after which the simulation was repeated to obtain the dynamic behaviour for the crack case.

3 Residual Signal Method

Early fault diagnosis is not always possible by only checking the trend of classical statistical features. For some systems, these statistical features are only able to react after a relatively large deviation of the trend.

Therefore, model-based methods have been developed to improve the fault diagnosis and to give a deeper insight into the system behaviour. These methods involve the generation of the residuals of the output variables indicating the difference between the healthy and the faulty cases [11, 12].

The model-based process can be divided into three steps; residual generation, residual evaluation and fault diagnosis [13]. In the current paper, three different techniques were applied for residual signal generation. The three techniques are namely; subtraction in the time domain, applying the comb filter in the frequency domain, and the auto-regressive AR model. A description of the three techniques can be found in Refs. [1–5]. These three techniques were applied using healthy and

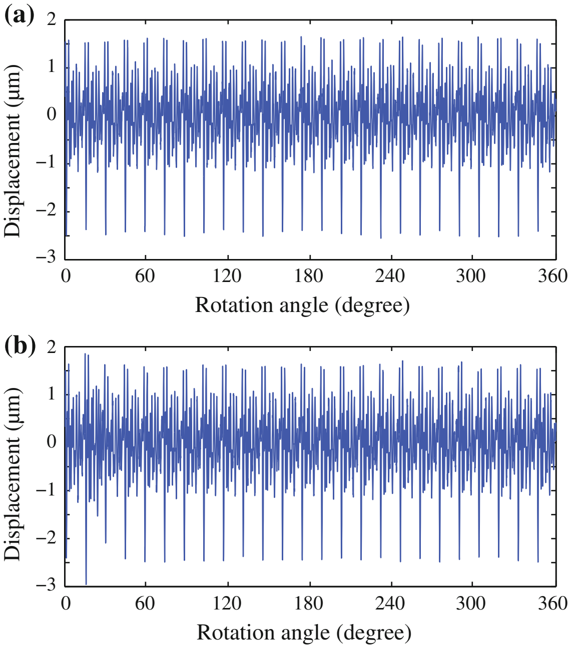
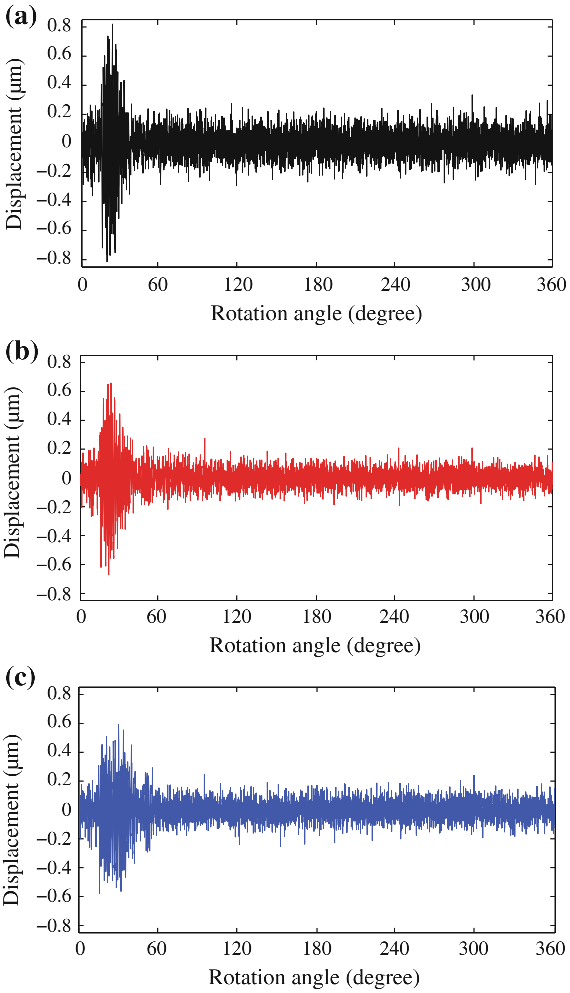


Fig. 3 Original signals obtained from dynamic simulation. **a** healthy case, **b** crack case with a 1 mm crack depth

Fig. 4 Residual signal generated using three different techniques. **a** subtraction in the time domain, **b** removing the gear mesh harmonics using the comb filter, **c** AR model using the Burg method with the order 200



faulty signals. The faulty signal was obtained from dynamic simulation for the case of a 1 mm crack depth, as well as the healthy signal, see Fig. 3. The results of the three applied techniques can be seen in Fig. 4.

4 Results and Discussion

The three residual signals shown in Fig. 4 are obtained for the same fault case. There are some differences which can be recognized. In the first technique which involves signal subtraction in the time domain, the peak indicating the impact of the crack is higher than those obtained with the two other techniques. This is because of the subtraction of two coincident signals representing the healthy and faulty signals. Moreover, because of the subtraction of the two random contents embedded in the healthy and faulty signals, the amount of the noise left in the residual signal is more than those obtained with the two other techniques, see Fig. 4a. To perform signal subtraction in the time domain, both the healthy and the faulty signals must start at exactly the same point of the same tooth to ensure the synchronisation of the time signals. This technique can be implemented in the time domain with simulated signals.

The second technique, involves removing the gear mesh frequencies using comb filter, has been applied, see Fig. 4b. The amount of the noise left in residual signal is less. The peak indicating the impact of the crack is obvious, but lower than that obtained from the first technique.

In Fig. 4c the result of the AR model is plotted. Based on the AR model used, the residual signal shows a relatively wider peak indicating the impact of the crack. AR technique is flexible in terms of different orders can be chosen for the prediction filter. Filter order should be carefully chosen in order to obtain a good prediction. High order filters can result in instability in prediction. In the current work the Burg method with the order 200 has been adopted.

5 Conclusions

The three residual techniques can be applied for residual signal generation. The first technique can be implemented in the time domain, but it requires a synchronisation of the two subtracted time signals. The amount of the noise with this technique is more than those obtained with the two other techniques. The two other techniques namely; using the comb filter in the frequency domain and the AR model, can be applied without the need of synchronised time signals.

References

1. Stewart RM (1977) Some useful data analysis techniques for gearbox diagnostics. In: Proceedings of meeting on application of time series analysis. ISVR, Southampton, UK, pp 18.1–18.19
2. Wang W, Wong AK (2002) Autoregressive model-based gear fault diagnosis. J Vib Acoust 124:172–179
3. Wu S (2007) Gearbox dynamic simulation and estimation of fault growth. MSc thesis, University of Alberta, Edmonton, Alberta, Canada
4. Wu S, Zuo M, Parey A (2008) Simulation of spur gear dynamics and estimation of fault growth. J Sound Vib 317(3–5):608–624
5. Mohammed OD, Rantatalo M, Aidanpaa J, Kumar U (2013) Vibration signal analysis for gear fault diagnosis with various crack progression scenarios. Mech Syst Signal Process 41: 176–195

6. Mahgoun H, Bekka RE, Felkaoui A (2012) Gearbox fault diagnosis using ensemble empirical mode decomposition (EEMD) and residual signal. Tribol Int 13(1):33–44
7. Chen Z, Shao Y (2011) Dynamic simulation of spur gear with tooth root crack propagating along tooth width and crack depth. Eng Fail Anal 18(8):2149–2164
8. Chaari F, Fakhfakh T, Haddar M (2009) Analytical modelling of spur gear tooth crack and influence on gearmesh stiffness. Eur J Mech A Solids 28(3):461–468
9. He S, Cho S, Singh R (2008) Prediction of dynamic friction forces in spur gears using alternate sliding friction formulations. J Sound Vib 309(3–5):843–851
10. He S, Gunda R, Singh R (2007) Effect of sliding friction on the dynamics of spur gear pair with realistic time-varying stiffness. J Sound Vib 301(3–5):927–949
11. Isermann R, Balle P (1997) Trends in the application of model-based fault detection and diagnosis of technical processes. Control Eng Pract 5:709–719
12. Isermann R (2005) Model-based fault-detection and diagnosis—status and applications. Annu Rev Control 29:71–85
13. Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 20:1483–1510

Prognostics and Structural Health Assessment Using Uncertain Measured Response Information

Achintya Haldar and Abdullah Al-Hussein

Abstract The authors and their team members have been working on developing implementable techniques for the objective rapid assessment of structural health (RASH) just after major natural and man-made events or in the context of maintenance over a period of time. They used the system-identification techniques by eliminating some of its weaknesses. For easier implementation, the excitation information was completely ignored. To locate defects and their severity at the local element level, the structures were represented by finite elements. By tracking the changes in the stiffness parameters of each element, the location(s) and severity of defects are assessed. The team conducted extensive analytical and laboratory investigations to verify all the methods. They had to overcome several challenges related to the conceptual and analytical development, data processing, and the presence of uncertainty in the every phase. To consider nonlinearity in the system identification process, a method known as Generalized Iterative Least Squares-Extended Kalman Filter-Unknown Input (GLIS-EKF-UI), was developed earlier. Since it failed to identify structures in some cases, the authors recently proposed a new method denoted as Unscented Kalman Filter—Unknown Input-Weighted Global Iterations (UKF-UI-WGI). With the help of informative examples, the superiority of UKF-UI-WGI over GLIS-EKF-UI is documented in this paper. Since at the beginning of an inspection, the defects and their severity are expected to be unknown, the authors recommend UKF-UI-WGI for the rapid assessment of health of infrastructures.

Keywords Structural health assessment · Uncertain measured information · Kalman filters · Nonlinear system identification · Unknown input excitation

A. Haldar (✉) · A. Al-Hussein
University of Arizona, 1209 E. 2nd St, Tucson, AZ 85721, USA
e-mail: haldar@u.arizona.edu

A. Al-Hussein
e-mail: abdullaa@email.arizona.edu

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_13

1 Introduction

ICRESH-ARMS 2015 provides a unique opportunity to discuss all the issues related to Prognostic and Structural Health Assessment. In fact, the first two issues of Life Cycle Reliability and Safety Engineering published by the Society for Reliability and Safety (SRESA) in 2015 are dedicated to the related topics. The related areas have become one of the most active research topics and have attracted multi-disciplinary interest. Extending life of infrastructures instead of replacing them has become a major challenge to engineers [9]. Structural health assessment just after a natural event or a man-made event has also become a part of inspection protocol. Non-destructive evaluation or inspection techniques of various degrees of sophistication are developed to help the assessment process. Smart sensing technologies, high quality data acquisition systems, mitigation techniques for noise contamination, digital communications, sophisticated computational techniques, etc., have been developed. This general area is commonly known as structural health assessment (SHA) or structural health monitoring (SHM).

Any automated monitoring practice that seeks to assess the health of a structure can be considered as SHM [7]. It implies that the health of a structure can be monitored in an automated manner by tracking the initiation or growth of a defect already present in the system. Since visual inspections may not be adequate for this purpose, sensors and the interpretation of their readings are essential for SHM. In spite of its recent impressive developments, it is not generally used in real world applications. Continuous accurate measurements of any output is a major challenge considering power sources necessary for operation, data transfer and storage, failure or sensors getting out of calibration, etc. The users generally assume that the technology is not fully developed for practical applications.

Objective rapid assessment of structural health (RASH) is essential just after a visual inspection or after a major natural event like strong earthquake or high wind or man-made event like blast or explosion, or in the context of maintenance. There is a potential for significant loss of economic activities in a region without such assessment. There are significant developments in the related areas. These areas are the subject of this paper.

2 Rapid Assessment of Structural Health

All defects are not equally important in maintaining the overall structural health. Thus, some of the major objectives of RASH are to locate defects at the local element level, assess their severity, and take remedial actions when necessary. If defects are repaired, it is important to know if they are repaired properly and all major defects are identified. To achieve these objectives, the process of listening to audible variations of responses due to tapping of structural surface has been used over centuries. Visual inspections at regular intervals are also suggested in many design codes. They can be broadly categorised as non-model based non-destructive inspection (NDI) techniques. If location of a defect is known, the profession now have technological sophistication to inspect it using instrument-based Penetrate Testing, Magnetic Particle Testing, Radiographic Testing, Ultrasonic Testing, Eddy Current Testing, Acoustic Emission Testing, Thermal Infrared Testing, etc.

For most large civil infrastructures, the location, number, and severity of defects may not be known in advanced. Sometimes, defects may be hidden behind obstructions like fire-proofing materials. Thus, instrument-based non-model approaches may not satisfied our needs. In the recent past, a consensus started developing about the use of measured time domain dynamic responses at the global level to assess the current structural health at the local element level. By appropriately tracking the signature embedded in the measurements, the structural health can be assessed.

The research team at the University of Arizona has been working on developing testing protocols for RASH for over two decades. After conducting extensive literature review, the team concluded that to locate defects, number, and their severity, it will be helpful if structures are represented by finite elements and their dynamic responses are measured in time domain representing their current state. By comparing the identified dynamic properties, essentially the stiffness properties of the elements, with the expected values, or reference values obtained from the design drawings, or changes from the previous values if inspections are carried out periodically, or variations from one member to another with similar sectional properties, the location(s), number, and severity of defects can be established at the element level. The concept is based on the axiom that the presence of defects will alter the dynamic responses and by tracking the signature embedded in the responses, the structural health can be assessed rapidly.

3 System Identification-Based Rash

By measuring dynamic excitation and response information, the stiffness parameter of all the elements in the finite element representation can be evaluated using an inverse mathematical concept commonly known as the system identification (SI) technique. However, Maybeck [21] correctly pointed out that deterministic mathematical models and control theories do not appropriately represent the behavior of a physical system and thus the SI-based method may not be appropriate. The research team successfully demonstrated that SI-based concept can be used for RASH if the different sources of uncertainty are accounted for appropriately and the system parameters are evaluated in an optimal sense using proper data processing algorithm.

3.1 System Identification with Unknown Input

One of the basic requirements for RASH is the simplicity in the inspection process. It is known to the profession that measuring dynamic excitation forces in the field condition can be very error prone due to inherent noises in the sensors and contamination due to multiple sources of excitation which is beyond the control of the inspector. It will be very desirable if a system can be identified using only measured response information completely ignoring the excitation information. The team developed several such techniques, commonly known as Iterative Least Squares with Unknown Input (ILS-UI) [24], Modified ILS-UI or MILS-UI [19], and Generalized ILS-UI or GILS-UI [18]. Mathematical concepts used to develop them cannot be discussed here due to lack of space but widely available in the literature. One major advantage of these procedures is that they are not very sensitive to the noises present in the response time histories.

3.2 System Identification with Unknown Input and Limited Response Information—Kalman Filter

One major deficiency of the methods discussed in the previous section is that they require response time histories at all DDOFs. To assess health of real infrastructures, it may be practically impossible and very expensive to install sensors at all DDOFs. In most cases, only a small part of the structure can be instrumented. When available responses are limited, generally Kalman filter (KF)-based concept is used. Kalman filter [15, 16] is a set of mathematical equations that provides efficient computational means in a recursive manner to estimate the state of a process, in a way that minimizes the mean of squared error, and calculates the best estimate of states from the noisy sensor responses [12, 26]. It is a time domain filter and is very powerful in several aspects. One of its limitations is that it is applicable for linear systems. If KF is used for RASH, the identification process becomes nonlinear. This is due to the fact that the identification of the unknown parameters jointly with dynamic responses is a nonlinear identification problem even if the structural system is linear. For nonlinear SI, extended Kalman filter (EKF) will be an attractive choice. It extends the linear Kalman filter to handle nonlinear systems based on a first-order linearization of the nonlinear statistical distributions of the variables. For RASH, EKF is an important requirement.

To implement EKF for RASH, the excitation force and the initial state vector must be known. The first requirement will defeat the purpose of SHA without input or ILS-UI. The second requirement is the final product of any inspection strategy and will not be available at the initiation of the inspection process. These two implementation requirements essentially limit the use of the basic KF concept for RASH.

Since EKF is very powerful, the team [25] decided to generate the required information to implement it. Suppose only a small part of the structure is instrumented. For the ease of discussion, it will be denoted as substructure. It is assumed that the responses at all DDOFs of the substructure will be measured. Then, the ILS-UI concept can be used to identify the stiffness parameter of all the elements in the substructure. All the beams and columns in the whole structure are expected to have similar cross sectional properties. Assuming the substructure contains a beam and a column element, all the elements in the whole structure can be assigned respective properties and the initial state vector of the structure will now be available. One very attractive attribute of ILS-UI is that it identifies the unknown excitation time history. Thus, with the introduction of the substructure concept, the two implementation requirements of EKF can be satisfied and the health of large real structural systems can be assessed using limited noise-contaminated responses without using any information on excitation.

The concept just discussed is known as Generalized Iterative Least Squares—Extended Kalman Filter—Unknown Input or GILS-EKF-UI. It can be implemented in two stages. In Stage 1, based on the available response information, a substructure can be identified. Using ILS-UI on the substructure, the unknown excitation time history and the stiffness parameter of all the elements in the substructure can be identified. The information will help to develop the initial state vector for the whole structure. Then in Stage 2, the EKF concept will be used to identify the stiffness parameter of all the elements in the structure. In this way, the number, location, and severity of defects can be assessed very accurately. The mathematical theories behind the two stages are discussed very briefly below.

4 Mathematics of Gils-Ekf-Ui

4.1 Stage 1—ILS-UI

The governing differential equation of motion using Rayleigh damping for the substructure can be expressed as:

$$\mathbf{M}_{sub}\ddot{\mathbf{X}}_{sub}(t) + (\alpha\mathbf{M}_{sub} + \beta\mathbf{K}_{sub})\dot{\mathbf{X}}_{sub}(t) + \mathbf{K}_{sub}\mathbf{X}_{sub}(t) = \mathbf{f}_{sub}(t) \qquad (1)$$

where $\mathbf{M}_{sub}$ is the global mass matrix, generally considered to be known; $\mathbf{K}_{sub}$ is the global stiffness matrix; $\ddot{\mathbf{X}}_{sub}(t)$, $\dot{\mathbf{X}}_{sub}(t)$, and $\mathbf{X}_{sub}(t)$ are the vectors containing the acceleration, velocity, and displacement, respectively, at time t; $\mathbf{f}_{sub}(t)$ is the input excitation vector at time t; and α and β are the mass and stiffness proportional Rayleigh damping coefficients, respectively. The subscript 'sub' is used to denote substructure.

The global mass and stiffness matrix can be formulated using standard procedure. The stiffness parameter for the ith element, k_i is defined as E_iI_i/L_i, where $L_{i,}$ I_i and E_i are the length, moment of inertia, and modulus of elasticity, respectively. The **P** vector contains all the unknown parameters and can be defined as:

$$\mathbf{P} = [k_1\ k_2\ \cdots\ k_{nesub}\ \beta k_1\ \beta k_2\ \cdots\ \beta k_{nesub}\ \alpha]^T \tag{2}$$

Using the least squares concept, it can be estimated as [24]:

$$\mathbf{P} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{F} \tag{3}$$

where **A** matrix contains the measured displacement and velocity responses at time point t; **F** vector contains the unknown input excitations and the inertia forces at time point t; and the responses are measured at equal interval of Δt for q time points. Since the input excitation $\mathbf{f}_{sub}$ is unknown, the force vector **F** in Eq. (3) is partially known and the iteration process cannot be initiated. To start the iteration process, the excitation information can be initially assumed to be zero for all the time points as discussed in [18]. The iteration process is continued until the excitation time history converges at all time points, considering two successive iterations, with a predetermined tolerance level. A tolerance level is set to be 10^{-8} in this study.

It is important to note that only acceleration time histories will be measured during an inspection. However, velocity and displacement time histories are necessary to implement the concept. The acceleration time histories can be successively integrated to generate the velocity and displacement time histories as discussed in more details in [8, 10, 22].

4.2 Stage 2—Implementation of EKF Concept

To implement the EKF concept, the differential equation in state-space form and the discrete time measurements can be expressed as:

$$\dot{\mathbf{Z}}(t) = f[\mathbf{Z}(t), t] \tag{4}$$

$$\mathbf{Y}(k) = h\,[\mathbf{Z}(k), t] + \mathbf{V}(k) \tag{5}$$

where $\mathbf{Z}(t)$ is the state vector at time t; $\dot{\mathbf{Z}}(t)$ is the time derivative of the state vector; f is a nonlinear function of the state; $\mathbf{Y}(k)$ is the measurement vector; h is the function that relates the state to the measurement; $\mathbf{V}(k)$ is a zero-mean, uncorrelated, white noise process with variance $R(k)$, and represented by $E[V(k)\ V^T(j)] = R(k)\delta(k-j)$, where $\delta(k-j)$ is the Kronecker delta function; that is $\delta(k-j) = 1$ if $k = j$, and $\delta(k-j) = 0$ if $k \neq j$.

For a structure represented by N number of degrees of freedom and L number of elements, the vectors $\mathbf{Z}(t)$ and $\dot{\mathbf{Z}}(t)$ are of size $(2\,N + L) \times 1$, L is the total number of unknown stiffness parameters. They are formed in the following way:

$$\mathbf{Z}(t) = \begin{bmatrix} \mathbf{Z}_1(t) \\ \mathbf{Z}_2(t) \\ \mathbf{Z}_3(t) \end{bmatrix} = \begin{bmatrix} \mathbf{X}(t) \\ \dot{\mathbf{X}}(t) \\ \tilde{\mathbf{K}} \end{bmatrix} \tag{6}$$

$$\dot{\mathbf{Z}}(t) = \begin{bmatrix} \dot{\mathbf{X}}(t) \\ \ddot{\mathbf{X}}(t) \\ 0 \end{bmatrix} = \begin{bmatrix} \dot{\mathbf{X}}(t) \\ -\mathbf{M}^{-1}[\mathbf{K}\mathbf{X}(t) + (\alpha\mathbf{M} + \beta\mathbf{K})\dot{\mathbf{X}}(t) - \mathbf{f}(t)] \\ 0 \end{bmatrix} \tag{7}$$

where $\tilde{\mathbf{K}} = [k_1 \quad k_2 \quad \cdots \quad k_{ne}]^T$ is column vector of size $(L \times 1)$.

For the identification of the whole structure, acceleration responses will be measured at a fewer (B) number of DDOFs. The accelerations will be integrated twice to obtain the velocities and displacements, as described in [22]. The vector $\mathbf{Y}(k)$ will have size $(2B \times 1)$ and will contain information on observed displacements and velocities.

Therefore, the discrete time measurement model is linear and it can be expressed at any discrete time k as:

$$\mathbf{Y}(k) = \mathbf{H} \cdot \mathbf{Z}(k) + \mathbf{V}(k) \tag{8}$$

where matrix $\mathbf{H}$ is the measurement matrix of size $2B \times (2\,N + L)$.

The filtering process in EKF can be started after initialization of state vector $\mathbf{Z}(0|0)$, which can be assumed to be Gaussian random variable with state mean $\hat{\mathbf{Z}}(0|0)$ and error covariance of $\mathbf{P}(0|0)$ i.e., $\mathbf{Z}(0|0) \sim N[\hat{\mathbf{Z}}(0), \mathbf{P}(0)]$.

The initial error covariance matrix $\mathbf{P}(0|0)$ contains information on the errors in the observed displacement and velocity responses, and in the initial values assigned to the unknown stiffness parameters of the whole structure. It is generally assumed to be diagonal and can be expressed as:

$$\mathbf{P}(0|0) = \begin{bmatrix} \mathbf{P}_x(0|0) & 0 \\ 0 & \mathbf{P}_s(0|0) \end{bmatrix} \tag{9}$$

where $\mathbf{P}_x(0|0)$ is a $(2\,N \times 2\,N)$ diagonal matrix, contains initial error covariance for observed responses; $\mathbf{P}_s(0|0)$ is a $(L \times L)$ diagonal matrix, contains initial error covariance for matrix $\tilde{\mathbf{K}}$. In the present study, a value of 1.0 is considered for the diagonal entries of $\mathbf{P}_x(0|0)$. Jazwinski [12] and Al-Hussein and Haldar [2, 4] pointed out that the diagonal entries for $\mathbf{P}_s(0|0)$ should be large positive numbers to accelerate the convergence of the local iteration process. A value of 1000 is used in this study.

The basic filtering process in EKF is the same Kalman filter (KF), i.e. propagation of the state mean and covariance from time k to one step forward in time

$k + 1$, and then updating them when the measurement at time $k + 1$ becomes available. Mathematically the steps can be expressed as:

(i) Prediction of state mean $\hat{\mathbf{Z}}(k+1|k)$ and its error covariance matrix $\hat{\mathbf{P}}(k+1|k)$ for the next time increment $k + 1$ as:

$$\hat{\mathbf{Z}}(k+1|k) = \hat{\mathbf{Z}}(k|k) + \int\limits_{k\Delta t}^{(k+1)\Delta t} \hat{\dot{\mathbf{Z}}}(t|k)dt \tag{10}$$

$$\mathbf{P}(k+1|k) = \mathbf{\Phi}(k+1|k)\mathbf{P}(k|k)\mathbf{\Phi}^T(k+1|k) \tag{11}$$

(ii) Using measurement $\mathbf{Y}(k+1)$ and Kalman gain $\mathbf{K}(k+1)$ available at time $k + 1$, updated state mean $\hat{\mathbf{Z}}(k+1|k+1)$ and error covariance matrix $\hat{\mathbf{P}}(k+1|k+1)$ can be obtained as:

$$\hat{\mathbf{Z}}(k+1|k+1) = \hat{\mathbf{Z}}(k+1|k) + \mathbf{K}(k+1)[\mathbf{Y}(k+1) - \mathbf{H}\cdot\hat{\mathbf{Z}}(k+1|k)] \tag{12}$$

$$\begin{aligned}\mathbf{P}(k+1|k+1) = {} & [\mathbf{I} - \mathbf{K}(k+1)\ \mathbf{H}]\ \mathbf{P}(k+1|k)\ [\mathbf{I} - \mathbf{K}(k+1)\ \mathbf{H}]^T \\ & + \mathbf{K}(k+1)\ \mathbf{R}(k+1)\ \mathbf{K}^T(k+1)\end{aligned} \tag{13}$$

where

$$\mathbf{K}(k+1) = \mathbf{P}(k+1|k)\mathbf{H}^T[\mathbf{H}\mathbf{P}(k+1|k)\mathbf{H}^T + \mathbf{R}(k+1)]^{-1} \tag{14}$$

where, $\mathbf{\Phi}(k+1|k)$ is the state transfer matrix from time k to $k + 1$; $\mathbf{K}(k+1)$ and $\mathbf{R}(k + 1)$ is the Kalman gain matrix and diagonal noise covariance matrix, respectively, at time $k + 1$. Detail procedure for calculation of $\mathbf{\Phi}$, $\mathbf{K}$, and $\mathbf{M}$ can be found in [17]. The symbol · stands for matrix multiplication. In the present study, diagonal entries in the noise covariance matrix $\mathbf{R}(k)$ are considered to be 10^{-4}.

5 UKF Based SI Concept

As will be discussed later, GILS-EKF-UI was successfully verified by conducting extensive analytical and laboratory investigations. In the laboratory investigations, the transverse acceleration time-histories were measured by capacitance accelerometers and angular rotation by autocollimators [20, 23]. To avoid contamination by other sources of excitations beyond the control of the inspector, responses were collected at a high sampling rate, 4000 cycles per second, for a fraction of a second. More recently it was observed that GILS-EKF-UI failed to converge or identify a structure when the sampling rate is much lower than what was used for the laboratory investigation. Upon further investigations, the authors concluded that the major reason for the non-convergence is the presence of higher level of nonlinearity. GILS-EKF-UI is supposed to identify a system in the presence of some degree of nonlinearity but the threshold is not known at this time. The first-order linearity used in EKF may not be sufficient to address more severe level of nonlinearities in the responses.

The authors [1] concluded that the unscented Kalman filter (UKF) concept can be used for highly nonlinear system identification problems. The UKF concept was developed by Julier et al. [14] to address the shortcoming of EKF. The UKF concept was developed based on unscented transformation (UT) with the underlying assumption that approximating a Gaussian distribution is easier than approximating a nonlinear transformation. UKF uses deterministic sampling to approximate the state distribution as a Gaussian Random Variable. The sigma points are chosen to capture the true mean and covariance of state distribution. They are propagated through the nonlinear system. UKF determines the mean and covariance accurately to the second order, while the EKF is only able to obtain first order accuracy [13].

The main difference between the EKF and UKF procedures is in the prediction step, i.e. prediction of the state vector and its error covariance using mathematical model of the system. They are the same in the updating step. In the prediction step of EKF, Jacobian matrices are used to linearize the nonlinear equations so that the linear KF can be used. However, in the prediction step of UKF, a number of state vectors or so-called sigma points is generated and then propagated through the nonlinear equations to get more accurate estimate. Thus, to implement the UKF procedure, instead of using Eqs. (10) and (11) of the EKF procedure, the following equations are necessary.

5.1 Sigma Points Calculation Step

At the current state vector $\hat{\mathbf{Z}}(k|k)$, sets of $2n + 1$ symmetric sigma points are generated so that they have the same mean and covariance of $\hat{\mathbf{Z}}(k|k)$ as following:

$$\begin{aligned}
\boldsymbol{\chi}_0(k|k) &= \hat{\mathbf{Z}}(k|k) \\
\boldsymbol{\chi}_i(k|k) &= \hat{\mathbf{Z}}(k|k) + \sqrt{(\lambda+n)}\mathbf{C}_{col,i} \quad i = 1,\ldots,n \\
\boldsymbol{\chi}_{i+n}(k|k) &= \hat{\mathbf{Z}}(k|k) - \sqrt{(\lambda+n)}\mathbf{C}_{col,i} \quad i = 1,\ldots,n
\end{aligned} \tag{15}$$

where

$$\lambda = \varphi^2(n+\gamma) - n \tag{16}$$

in which $\mathbf{C}$ is a square root of the covariance matrix such that $\mathbf{P}(k) = \mathbf{C} \cdot \mathbf{C}^T$; $\mathbf{C}_{col,i}$ is the ith column of $\mathbf{C}$'s matrix; n is the dimension of the state vector ($n = 2\,N + L$); The parameter φ determines the spread of the sigma points around the mean. Typical range value for φ is ($0 \leq \varphi \leq 1$). The parameter γ is a tertiary scaling factor and is usually set equal to 0. In fact, parameter γ can be used to reduce the higher order errors of the mean and the covariance approximations. Note that sigma points are a set of vectors whose components are real numbers.

5.2 Prediction Step

The sigma points are propagated through the nonlinear dynamic equation as:

$$\boldsymbol{\chi}_i(k+1|k) = \boldsymbol{\chi}_i(k|k) + \int_{k\Delta t}^{(k+1)\Delta t} f[\mathbf{Z}(t), t]dt \quad i = 1, \ldots, 2n \tag{17}$$

The predicted state vector $\hat{\mathbf{Z}}(k+1|k)$ can be shown to be:

$$\hat{\mathbf{Z}}(k+1|k) = \sum_{i=0}^{2n} W_i\ \boldsymbol{\chi}_i(k+1|k) \tag{18}$$

and its predicted error covariance matrix $\mathbf{P}(k+1|k)$ can be expressed as:

$$\begin{aligned}\mathbf{P}(k+1|k) = \sum_{i=0}^{2n} W_i \left[\boldsymbol{\chi}_i(k+1|k) - \hat{\mathbf{Z}}(k+1|k)\right]\left[\boldsymbol{\chi}_i(k+1|k) - \hat{\mathbf{Z}}(k+1|k)\right]^T \\ (1 - \varphi^2 + \psi)\left[\boldsymbol{\chi}_0(k+1|k) - \hat{\mathbf{Z}}(k+1|k)\right]\left[\boldsymbol{\chi}_0(k+1|k) - \hat{\mathbf{Z}}(k+1|k)\right]^T\end{aligned} \tag{19}$$

where ψ is the secondary scaling factor used to emphasize the weighting on the zero's sigma point for the covariance calculation. The value of ψ is greater than 0 and the best value is 2 for Gaussian distribution. The weight factor W_i can be shown to be:

$$W_0 = \frac{\lambda}{\lambda + n} \qquad i = 0 \tag{20}$$

$$W_i = \frac{1}{2(\lambda + n)} \quad i = 1, \ldots, 2n \tag{21}$$

It is important to point out here that in this study the measurement model is linear and linear KF is used to predict the measurement vector and its error covariance matrix.

5.3 Improvements in UKF Algorithm

When the EKF concept was used in the context of ILS-UI, i.e. the two-stage concept used in GILS-EKF-UI, it failed to converge in some cases. The authors observed that the use of UKF to identify large structural systems were very limited in the literature. Most of the reported works were developed to identify shear-type structures with very few DDOFs using long duration responses in one global iteration. Suppose that the responses are available for q time points. The iteration processes between successive time points in the UKF procedure are termed as local iterations and the iteration processes for all q time points are termed as a global iteration. The three steps of the UKF (sigma point, prediction and updating operations) are carried out for all q time points.

To obtain optimal, stable, and convergent solutions of the SI process, the authors proposed to use several global iterations using responses collected for a fraction of second. They noted that the error covariance matrix of the stiffness parameters reduced significantly during the successive global iterations and the identified stiffness values sometimes converge to the wrong values particularly when the initial values are far from the expected values representing defective states. This prompted the authors [3, 4] to introduce a weighted global iteration factor, w, to the error covariance matrix after the first global iteration so that the algorithm can detect the stiffness parameters with incorrect initial value but converges to the correct solution. In the second global iteration, the initial values of the stiffness parameters are the same as that of obtained at the completion of first global iteration. A weight factor w is introduced in the stiffness covariance matrix obtained at the completion of the first global iteration to amplify it and then used it as the initial stiffness covariance in the second global iteration. The weighted global iteration concept can be mathematically presented as:

$$\hat{\mathbf{Z}}^{(2)}(0|0) = \begin{bmatrix} \hat{\mathbf{X}}^{(2)}(0|0) \\ \hat{\dot{\mathbf{X}}}^{(2)}(0|0) \\ \tilde{\mathbf{K}}^{(2)}(0|0) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{X}}^{(1)}(0|0) \\ \hat{\dot{\mathbf{X}}}^{(1)}(0|0) \\ \tilde{\mathbf{K}}^{(1)}(q|q) \end{bmatrix} \tag{22}$$

$$\mathbf{P}^{(2)}(0|0) = \begin{bmatrix} \mathbf{P}_x^{(1)}(0|0) & 0 \\ 0 & w\mathbf{P}_s^{(1)}(q|q) \end{bmatrix} \tag{23}$$

The same processes of local iterations are carried out for all the time points and a new set of state vector and error covariance matrix are obtained at the completion of second global iteration. The weighted global iteration processes are continued until the estimated error in the identified stiffness parameters at the end of two consecutive global iterations becomes smaller than a predetermined convergence criterion (ε_s).

$$\left|\tilde{\mathbf{K}}^{(i)}(q|q) - \tilde{\mathbf{K}}^{(i-1)}(q|q)\right| \leq \varepsilon_s \times \left|\tilde{\mathbf{K}}^{(i-1)}(q|q)\right| \tag{24}$$

where i represents the ith global iteration. ε_s is considered to be 1 % in this study.

Although the weighted global iterations play an important role in the later stage to assure convergence; the global iteration procedure does not guarantee the convergence of the iteration scheme. If they diverge, the best estimated values based on minimum objective function $\bar{\theta}$ are considered, as discussed in [4, 11].

The procedure developed this way will be denoted as Unscented Kalman Filter —Unknown Input- Weighted Global Iterations or UKF-UI-WGI. It will be implemented in two stages in the same way as that of GILS-EKF-UI. It will not require any additional resources but it will improve the defect detection capability in a significant way, as will be elaborated further with the help of several informative examples.

6 Examples

It is hoped that the sequential development processes used by the research team to develop several RASHs for infrastructures are informative. However, during each phase of the development, the reviewers of technical papers commented that the procedures were reasonable from the theoretical point of view but could not be used for the health assessment of real infrastructures. This prompted the research team to initiate several laboratory investigations. One of them is discussed briefly below.

6.1 Example 1

6.1.1 Description of the Frame and Dynamic Testing

A two-dimensional one-bay three-story steel frame, shown pictorially in Fig. 1, was initially tested to verify the EKF procedure [20]. To fit the testing facilities, the frame was scaled to one-third of its actual dimensions. The scaled frame has a bay width of 3.05 m and story height of 1.22 m. The frame consists of nine members; six columns and three beams. Steel section of size S4 × 7.7 was used for all the beams and columns in order to minimize the effects of fabrication defects and differences in material properties. The frame was reconfigurable, i.e. bolted joints were used so that the defect-free and defective members could be interchanged to study defect detection capability. Several types of defects, very severe to minor in

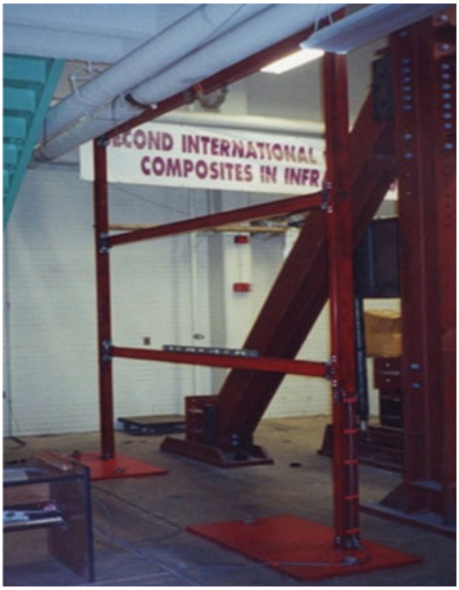


Fig. 1 The pictorial view of the frame

nature, were introduced. Some of the defect scenarios considered were removing a member completely, loss of area of a member over a finite length, multiple cracks in a member, one crack in a member, loosening bolts at joints, and multiple combinations of these defects. The same response information was used to verify both GILS-EKF-UI and UKF-UI-WGI in the following sections.

The frame consists of 9 members; 3 beams and 6 columns. The frame is represented by the finite element (FE) with 9 elements and 8 nodes. Each node has three DDOFs; two translational and one rotational. The support condition at the bases is considered to be fixed. Therefore, the total number of DDOFs for the frame is 18. The actual stiffness parameters k_i, defined in terms of (E_iI_i/L_i), for the beam and column are estimated to be 96500 and 241250 N-m, respectively. The first two natural frequencies of the defect-free frame were estimated experimentally to be $f_1 = 9.76$ Hz and $f_2 = 34.12$ Hz. Then, assuming the same damping for the first two significant frequencies, a procedure suggested in [6], is used to calculate the Rayleigh damping coefficients α and β. They are found to be 1.1453681 and 0.0000871, respectively. The frame is excited by a sinusoidal load $f(\mathrm{t}) = 1.4 \sin(58.23\mathrm{t})$ N applied at node 1, as shown in Fig. 2. Before conducting any test, numerous analytical verifications were carried under various testing conditions. For the analytical verifications, the responses of the frame in terms of displacement, velocity and acceleration time histories were numerically generated using a commercial software ANSYS (ver. 15.0) [5] at all 9 DDOFs (responses at nodes 1, 2 and 3) of the substructure for all cases. The frame is identified using responses from 0.02 to 0.32 s with time increment of 0.00025 s providing a total of 1201 time points. For the laboratory investigation, the translational and rotational acceleration

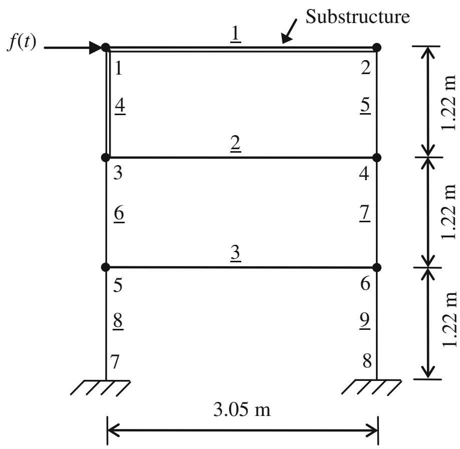


Fig. 2 Finite element representation of the test frame

Table 1 Stiffness parameter (*EI/L*) identification for the substructure—defect-free frame

Member	Nominal (N-m)	Identified	Change (%)
(1)	(2)	(3)	(4)
k_1	96500	96502	0.002
k_4	241250	241255	0.002

time histories were measured. They were successively integrated to generate velocity and displacement time histories as suggested in [8, 22].

6.1.2 Identification of the Defect-Free State of the Frame

To implement both the GILS-EKF-UI and UKF-UI-WGI methods, the substructure used is shown in double lines in Fig. 2. The stiffness parameters of the two elements in the substructure using ILS-UI in Stage 1 are identified and the results are summarized in Table 1. The results indicate that the substructure is identified very accurately. As mentioned earlier, ILS-UI also identifies the unknown excitation force. Both the actual and identified excitation time histories are shown in Fig. 3. The figure clearly indicates the unknown excitation time history is also identified very accurately.

The errors in measurement noises (*R)* in Eq. 5 are one of the important factors that influence the identification of the stiffness parameter. Two different values of *R* (10^{-3} and 10^{-4}) are considered in this study. Using the information from Stage 1, the stiffness parameter of all the nine members of the whole frame is identified using the GILS-EKF-UI and UKF-UI-WGI methods. The results are summarized in Table 2. As commonly used in the literature, the errors are defined as the percentage deviation of identified values, representing the current state, with respect to the initial theoretical values. The maximum acceptable error in the identification is

Fig. 3 Actual and identified force time histories using ILS-UI for defect-free case

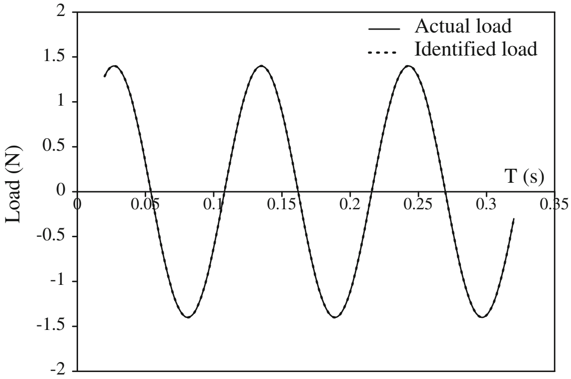


Table 2 Stiffness parameter (*EI/L*) identification for defect-free frame

Member	Nominal (N-m)	Error in Identification (%)			
		R = 10^{-4}		R = 10^{-3}	
		EKF	UKF	EKF	UKF
(1)	(2)	(3)	(4)	(5)	(6)
k_1	96500	0.002	−0.069	0.000	−0.030
k_2	96500	0.062	0.091	0.064	0.054
k_3	96500	0.047	0.096	−0.102	−0.065
k_4	241250	−0.063	−0.063	−0.004	−0.007
k_5	241250	−0.237	−0.073	−0.063	−0.001
k_6	241250	−0.096	0.013	−0.003	0.009
k_7	241250	−0.338	−0.104	−0.011	−0.015
k_8	241250	−0.032	−0.222	−0.011	−0.040
k_9	241250	−0.105	−0.206	−0.016	−0.039

about 10 % reported in the literature [2]. The results in Table 2 clearly indicate that both methods identified the stiffness parameters of all the members reasonably well for both measurement errors. In an overall sense, UKF-UI-WGI identified the frame more accurately than GILS-EKF-UI. Since the differences in identified stiffness parameters are relatively small, the health of the frame can be considered as defect-free.

6.1.3 Health Assessment of Defective Frame

After successfully identifying the defect-free frame, several defective states of the frame were considered, as discussed earlier. Only two defect scenarios are presented in the following sections.

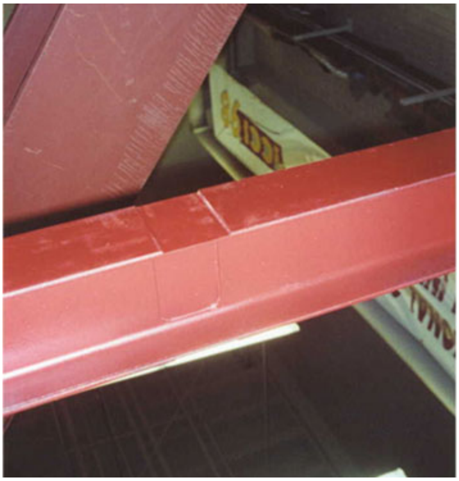


Fig. 4 Defect in member 3

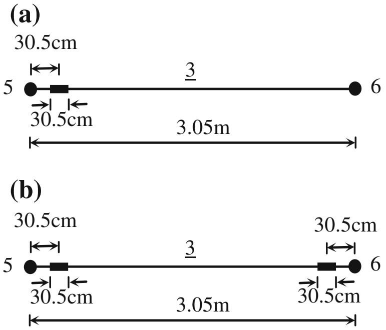


Fig. 5 Defects in the frame

Defect 1

In defect 1, member 3, the beam at the first story level, is considered to have one defect. The cross-sectional area of member 3 is considered to be corroded over a length of 30.5 cm, located at a distance of 30.5 cm from node 5. It is pictorially shown in Fig. 4. The defect is shown in Fig. 5a in the finite element representation.

The web and flange thicknesses are considered to be reduced by 20 % of their original values. The loss of thicknesses will result in the reduction of the cross-sectional area by 19.13 % and the moment of inertia by 17.02 % from the defect-free case. The identified stiffness parameters for all nine members using the GILS-EKF-UI and UKF-UI-WGI methods are summarized in Table 3. In all cases, the maximum changes occur in member 3, indicating it contains the defect. The results also indicate that both methods can be used for RASH of the frame.

Defect 2

In defect case 2, member 3 is considered to have two defects. The first defect is the same as that in defect case 1. For the second defect, the cross-sectional area is also

Table 3 Stiffness parameter (*EI/L*) identification for defect 1

Member	Nominal (N-m)	Change in Identification (%)			
		R = 10^{-4}		R = 10^{-3}	
		EKF	UKF	EKF	UKF
(1)	(2)	(3)	(4)	(5)	(6)
k_1	96500	−0.008	−0.107	0.048	0.037
k_2	96500	−0.023	0.115	−0.481	−0.314
k_3	96500	**−2.551**	**−2.609**	**−2.371**	**−2.472**
k_4	241250	−0.057	−0.009	−0.015	0.000
k_5	241250	−0.366	−0.148	−0.040	−0.024
k_6	241250	−0.211	−0.158	−0.091	−0.104
k_7	241250	−0.398	−0.099	−0.062	−0.058
k_8	241250	−0.239	−0.402	−0.192	−0.229
k_9	241250	−0.321	−0.411	−0.200	−0.237

Table 4 Stiffness parameter (*EI/L*) identification for defect 2

Member	Nominal (N-m)	Change in Identification (%)			
		R = 10^{-4}		R = 10^{-3}	
		EKF	UKF	EKF	UKF
(1)	(2)	(3)	(4)	(5)	(6)
k_1	96500	0.131	0.028	0.061	0.022
k_2	96500	−0.347	−0.159	−0.199	−0.141
k_3	96500	**−4.997**	**−5.088**	**−5.189**	**−5.055**
k_4	241250	−0.071	−0.025	−0.091	0.004
k_5	241250	−0.222	−0.020	−0.329	0.016
k_6	241250	−0.223	−0.193	−0.069	−0.055
k_7	241250	−0.467	−0.176	−0.223	−0.192
k_8	241250	−0.440	−0.594	−0.421	−0.653
k_9	241250	−0.508	−0.599	−0.449	−0.661

considered to be corroded over a length of 30.5 cm but it is located at a distance of 30.5 cm from node 6, as shown in Fig. 5b. The identified stiffness parameters for all members using the GILS-EKF-UI and UKF-UI-WGI methods are summarized in Table 4. In all cases, the maximum changes occur in member 3, indicating it contains the defect. The reduction in the stiffness parameter of member 3 for defect 2 is more than that of defect case 1. It is clearly indicated that the defect in case 2 is more severe than that in case 1. The results also indicate that both methods can be used for RASH of the frame.

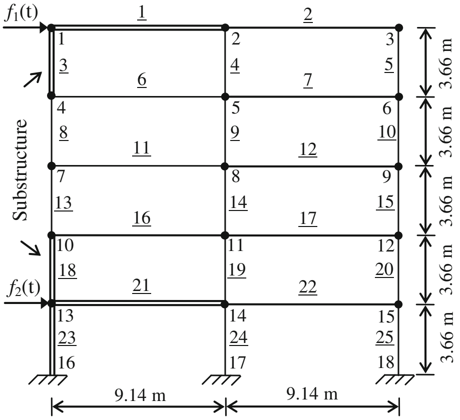


Fig. 6 Finite element representation of a frame

6.2 Example 2

In Example 1, both the GILS-EKF-UI and UKF-UI-WGI methods appear to identify the defect spot and the severity accurately. To demonstrate the superiority of UKF-UI-WGI over GILS-EKF-UI, this second example is considered.

6.2.1 Description of the Frame

A two-dimensional frame with a bay width of 9.14 m and story height of 3.66 m, as shown in Fig. 6, is considered. The frame has a total of 25 members; 10 beams and 15 columns. The beams and columns are made of W21 × 68 and W14 × 61 sections, respectively, of Grade 50 steel. The frame is modeled by 18 nodes in the finite element (FE) representation. Each node has three dynamic degrees of freedom (DDOFs); two translational and one rotational. The support condition at the base (nodes 16, 17, and 18) of the frame is considered to be fixed. The total number of DDOFs for the frame is 45. The actual theoretical stiffness parameter values k_i evaluated in terms of (E_iI_i/L_i) are calculated to be 13476 kN-m and 14553 kN-m for a typical beam and column, respectively. First two natural frequencies of the frame are estimated to be f_1 = 3.598 Hz and f_2 = 11.231 Hz, respectively. Following the procedure described in [6], Rayleigh damping coefficient α and β are calculated to be 1.7122088 and 0.00107326, respectively, for an equivalent modal damping of 5 % (commonly used in model codes in the US) of the critical for the first two modes.

The frame is excited simultaneously by two sinusoidal loadings. The first loading, $f_1(t)$ = 3 sin(18t) kN is applied horizontally at node 1, and the second loading, $f_2(t)$ = 2 sin(22t) kN is applied horizontally at node 13, as shown in Fig. 6. For this example, the information on responses are numerically generated using a commercially available software ANSYS (ver. 15.0) [5]. The responses are obtained at 0.0001 s time interval. After the responses are simulated, the information on input excitations is completely ignored. Responses between 0.02 and 0.32 s providing 3001 time points are used in the subsequent health assessment process.

6.2.2 Identification of the State of the Frame

Two substructures are considered to assess the health of this large frame. They are shown in Fig. 6 with double lines. Using responses at 18 DDOFs in the substructures, the stiffness and damping parameters and the time history of unknown input force are identified using the ILS-UI procedure in Stage 1, initially for the defect-free state of the frame. The errors in identification of the stiffness parameters are shown in Table 5. From the results, it can be observed that the errors in the identified stiffness parameter of the five members in the substructures are very small. The damping coefficients and excitation time history are also identified very accurately.

The information of Stage 1 is used to initiate both the GILS-EKF-UI and UKF-UI-WGI procedures. Then, the stiffness parameters of all 25 elements of the frame are estimated. The stiffness parameters of all members in the frame are identified for the defect-free state and the results are summarized in Table 6, Columns 3 and 4, respectively, for both methods. Since the identified stiffness parameter did not vary significantly from the expected values, the methods correctly identified the defect-free state of the frame. The results of GILS-EKF-UI are still within the acceptable level but not as good as the UKF-UI-WGI method. However, it can be concluded that both filters identified the defect-free state of the frame.

After assessing structural health of the defect-free frame, one defective case is considered for this example. In Defect 1, the cross-sectional area of member 17 is considered to be corroded over a length of 30 cm, located at a distance of 30 cm from node 12. The results for the substructure identification in Stage 1 using ILS-UI are summarized in Table 5, Columns 5 and 6. As for the defect-free case, for this defective state, the substructures are identified accurately. Using the information from Stage 1, the whole frame is then identified using both methods in Stage 2. The

Table 5 Stiffness parameter (*EI/L*) identification of the substructure for Example 2

Member	Theoretical (kN-m)	Defect Free		Defect 1	
		Identified	Change (%)	Identified	Change (%)
(1)	(2)	(3)	(4)	(5)	(6)
k_1	13476	13476	0.001	13476	0.001
k_3	14553	14553	0.001	14553	0.001
k_{18}	14553	14553	0.004	14553	0.003
k_{21}	13476	13477	0.003	13477	0.003
k_{23}	14553	14553	0.004	14553	0.003

Table 6 Change (%) in stiffness parameter (*EI/L*) identification of whole structure

Member	Theoretical (kN-m)	Defect Free		Defect 1	
		EKF	UKF	EKF	UKF
(1)	(2)	(3)	(4)	(5)	(6)
k_1	13476	−0.05	−0.07	−0.07	−0.06
k_2	13476	−0.37	0.37	−6.84	0.87
k_3	14553	−0.03	−0.05	−0.07	−0.05
k_4	14553	0.03	0.17	−1.96	0.20
k_5	14553	0.76	−0.13	11.23	−1.00
k_6	13476	−0.06	−0.02	0.73	−0.02
k_7	13476	−0.04	−0.21	1.90	0.06
k_8	14553	0.41	0.67	−0.12	0.62
k_9	14553	0.38	0.57	−2.88	0.37
k_{10}	14553	0.69	0.22	7.31	1.37
k_{11}	13476	0.51	0.09	2.43	0.25
k_{12}	13476	−0.25	0.01	−2.06	−0.71
k_{13}	14553	−0.68	−0.55	−1.07	−0.57
k_{14}	14553	−1.57	−0.81	−4.68	−1.69
k_{15}	14553	0.07	−1.17	4.42	−0.13
k_{16}	13476	0.40	0.26	0.56	−0.09
k_{17}	13476	1.24	1.01	**−7.54**	**−8.39**
k_{18}	14553	0.01	0.02	0.17	0.05
k_{19}	14553	−0.69	−0.45	−1.07	−0.42
k_{20}	14553	0.26	0.02	−1.25	−1.16
k_{21}	13476	0.07	0.04	0.18	0.05
k_{22}	13476	−0.52	−0.45	−1.34	−0.97
k_{23}	14553	0.14	0.06	0.18	0.05
k_{24}	14553	0.09	0.00	0.02	0.05
k_{25}	14553	0.15	0.34	0.67	0.62

results in Columns 5 and 6 in Table 6 clearly indicate that the UKF-UI-WGI procedure is capable of identifying the location and severity of defect. The identification of defect location using the GILS-EKF-UI procedure for the defective case is not straightforward. Both the UKF and EKF-based procedures identified the reductions of the stiffness parameter of defective member 17 as 8.39 and 7.54 %, respectively. However, the results of EKF-based procedure show that the stiffness parameter of defect-free member 5 is increased by 11.23 %, which is more than acceptable error. Therefore, it can be concluded that GILS-EKF-UI failed to assess the health of the frame for the defective state. This example clearly demonstrates the superiority of the proposed UKF-UI-WGI procedure over the GILS-EKF-UI procedure developed earlier by the research team.

7 Conclusions

The rapid assessment of structural health has become a major challenge in the context of routine maintenance or just after major natural and man-made events. The authors and their team members used the system-identification techniques by mitigating its weaknesses to identify defects and their severity at the local element level by representing real structures using finite elements. For easier implementation, the excitation information was completely ignored. By tracking the changes in the stiffness parameters of each element the location(s) and severity of defects are assessed. The team conducted extensive analytical and laboratory investigations to verify all the methods. They had to overcome several challenges related to the conceptual and analytical development, data processing, and the presence of uncertainty in the every phase. To consider nonlinearity in the system identification process, a method known as Generalized Iterative Least Squares-Extended Kalman Filter-Unknown Input (GILS-EKF-UI), was developed by the team earlier. Since it failed to identify structures in some cases, the authors recently proposed a new method denoted as Unscented Kalman Filter—Unknown Input- Weighted Global Iterations (UKF-UI-WGI). With the help of informative examples, the superiority of UKF-UI-WGI over GILS-EKF-UI is documented in this paper. Since at the beginning of an inspection, the defects and their severity are expected to be unknown, the authors recommend UKF-UI-WGI instead of GILS-EKF-UI for the rapid assessment of health of infrastructures.

Acknowledgments The authors would like to thank all the team members for their help in developing the overall research concept of structural health assessment and monitoring. They include Drs. Duan Wang, Peter H. Vo, Xiaolin Ling, Hasan N. Katkhuda, Rene Martinez-Flores, Ajoy K. Das, Mr. J. P. Kazakoff and A. R. Safdar. The team received financial supports from the National Science Foundation, a small grant from the University of Arizona, graduate student supports from various sources including Raytheon Missile Corporation, Ministry of Higher Education, Jordon, CONACYT, Iraq's Ministry of Higher Education and Scientific Research, Department of Civil Engineering and Engineering Mechanics, University of Arizona, etc. The financial supports from the National Science Foundation; most recently under Grant No. CMMI-1403844 is also appreciated. Any opinions, findings, or recommendations expressed in this paper are those of the writers and do not necessarily reflect the views of the sponsors.

References

1. Al-Hussein A, Haldar A (2015) A comparison of unscented and extended Kalman filtering for nonlinear system identification. In: 12th international conference on applications of statistics and probability in civil engineering (ICASP12-2015). Vancouver, Canada
2. Al-Hussein A, Haldar A (2015) Novel unscented Kalman filter for health assessment of structural systems with unknown input. J Eng Mech ASCE 141(7):04015012-1–04015012-13. doi:10.1061/(ASCE)EM.1943-7889.0000926
3. Al-Hussein A, Haldar A (2015) Structural health assessment at a local level using minimum information. Eng Struct 88:100–110. doi:10.1016/j.engstruct.2015.01.026

4. Al-Hussein A, Haldar A (2015) Unscented Kalman filter with unknown input and weighted global iteration for health assessment of large structural systems. Structural Control and Health Monitoring. doi:10.1002/stc.1764
5. ANSYS version 15.0 (2013).The Engineering Solutions Company
6. Clough RW, Penzien J (2003) Dynamics of structures, 3rd edn. Computers and Structures, California
7. Cross EJ, Worden K, Farrar CR (2013) Chapter 1—structural health assessment for civil infrastructure. Health Assessment of Engineered Structures: Bridges, Buildings and Other Infrastructures, World Scientific Publishing Co., Editor, pp 1–31
8. Das AK (2012) Health assessment of three dimensional large structural systems using limited uncertain dynamic response information. PhD dissertation, Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, Arizona
9. Haldar A (ed) (2013) Health assessment of engineered structures: bridges, buildings and other infrastructures. World Scientific Publishing Co., New Jersey
10. Haldar A, Das AK, Al-Hussein A (2013) Data analysis challenges in structural health assessment using measured dynamic responses. Adv Adapt Data Anal 5(4):1350017-1–1350017-22
11. Hoshiya M, Saito E (1984) Structural identification by extended Kalman filter. J Eng Mech ASCE 110(12):1757–1772
12. Jazwinski AH (1970) Stochastic process and filtering theory. Academic Press, Inc., New York
13. Julier SJ, Uhlmann JK (2004) Unscented filtering and nonlinear estimation. In: Proceedings of the IEEE, vol 92, no 3, pp 401–422
14. Julier SJ, Uhlmann JK, Durrant-Whyte HF (1995) A new approach for filtering nonlinear systems. In: Proceedings of the american control conference, vol 3. Seattle, Washington, pp 1628–1632
15. Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82:35–45
16. Kalman RE, Bucy RS (1961) New results in linear filtering and prediction theory. J Fluids Eng 83(1):95–108
17. Katkhuda H, Haldar A (2008) A novel health assessment technique with minimum information. Struct Control Health Monit 15(6):821–838
18. Katkhuda H, Martinez-Flores R, Haldar A (2005) Health assessment at local level with unknown input excitation. J Struct Eng ASCE 131(6):956–965
19. Ling X, Haldar A (2004) Element level system identification with unknown input with Rayleigh damping. J Eng Mech ASCE 130(8):877–885
20. Martinez-Flores R (2005) Damage assessment potential of a novel system identification technique—experimental verification. PhD dissertation, Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, Arizona
21. Maybeck PS (1979) Stochastic models, estimation, and control theory. Academic Press Inc, UK
22. Vo PH, Haldar A (2003) Post-processing of linear accelerometer data in structural identification. J Struct Eng 30(2):123–130
23. Vo PH, Haldar A (2008) Health assessment of beams—experimental verification. Struct Infrastruct Eng 4(1):45–56
24. Wang D, Haldar A (1994) Element-level system identification with unknown input information. J Eng Mech ASCE 120(1):159–176
25. Wang D, Haldar A (1997) System identification with limited observations and without input. J Eng Mech ASCE 123(5):504–511
26. Welch G, Bishop G (1995) An introduction to the Kalman filter. Technical report TR95-041, Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Ontology Based Diagnosis for Maintenance Decisions of Paper Mill Roller Using Dynamic Response

Madhav Mishra and Adithya Thaduri

Abstract Context-aware systems have been applied in several fields like Information Technology, mobile, web services, travel guidance etc. These systems deliver decisions based on a 'context' by using contextual models. In paper industries, the failures of rollers were prominent and rolling element bearing is one of the critical components. The failure occurs due to the varying levels of the loads and external parameters that defines context. This paper demonstrates the ontology contextual modeling for the diagnosis of rollers as a context by using dynamic response. The roller is modeled using physical models and applying runs of different parameters and its levels. Then contextual models are generated for rollers to show relation among input contextual parameters with different features. This paper shows that this conceptual idea of decision based on different contexts using ontology models is for effective diagnosis facilitate maintenance strategies and further prospects in prognosis.

Keywords Paper mill roller · Context-aware systems · Dynamic response · Relational models · Machine learning · Diagnosis

1 Introduction

With the rapid expansion of scientific technology, the machines used in modern industries are becoming larger, precise increasingly automated. Their structures become more multifaceted and their probable faults become more challenging to find. So, in the field of mechanical fault diagnosis, it is an urgent problem to exactly evaluate and correctly predict the running condition of the mechanical

M. Mishra (✉) · A. Thaduri
Luleå University of Technology, 97187 Luleå, Sweden
e-mail: madhav.mishra@ltu.se

A. Thaduri
e-mail: adithya.thaduri@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_14

equipment [1]. Roller bearing is one of the utmost extensively used elements in rotary machines. In most machinery, bearing is one of the essential components that directly influence the operation of the whole machinery. Faulty bearings cause majority of problems in the rotary machinery [2]. Localized defect is the main failure mode of rolling element bearings. Vibration and acoustic emission (AE) signals are widely used in condition monitoring of rotating machines.

During operating process, the machine set can generate all kinds of signals, and involve in many correlated features. When any of these features deviates beyond their specified limits, a fault may emerge. How to effectively extract the fault feature, which can correctly reflect the occurrence of the fault, is still an ongoing research issue [3]. In rotating machinery, the failure of rolling element bearing can result in the deterioration of machine running condition. Effectively detecting and diagnosing the incipient fault of rolling element bearing can provide an assurance for the reliability of machine set running. Generally, extracting the fault feature from the vibration signal to detecting the occurrence of the fault can effectively reduce the possibility of catastrophic damage and the downtime [4]. Therefore, quite naturally, fault identification of rolling element bearings has been the subject of extensive research [5]. Fault detection is possible by comparing the signals of a machine running in normal and faulty conditions. The faults considered in the present study are inner race fault (IRF), outer race fault (ORF) and inner and outer race fault (IORF) [6]. Machine condition monitoring system is a decision support tool, which is capable of identifying the failure of a machine and capable of predicting failure from its symptoms [7].

In the case of paper industries, the rolling element bearing in the rollers are very important because the failure of these bearings results in stoppage of production [8]. To ensure operation of the paper mill rolling element bearing, vibration condition monitoring techniques are implemented in this field to identify, isolate and mitigate the failures [9]. Because the paper mill runs continuously, a certain amount of paper dust on the felt wire will act as an extra load on the rollers. To reduce the load, a particular maintenance action is followed to remove the dust by regular intervals. Normally, this action is done at regular intervals without the use of condition monitoring. Apart from this dust, there are other maintenance actions that can also be applied on the roller without any intelligence. The main problem of this maintenance actions are the irregular work stoppages, corrosion by lubricant quality, inactive human skill, cleaning, operating costs and unawareness of the surrounding environment [10]. The main objective of this work is to establish a decision support mechanism that provides necessary actions on maintenance depends on the condition monitoring and operating environment.

There were studies that used dynamic response for the rolling element bearings for non-linearity [11], for stiffness [12] and for transient rotor dynamics [13]. The modelling for diagnosis by physical modelling of rolling element bearings were carried out by using support vector machine (SVM) [14], wavelet packets [15] and neural network approach [16]. There have been efforts to provide these decisions on rolling element bearing for diagnosis. The classification performance of various fractal dimensions and their combinations on different fault data sets were studied on rolling element bearing using support vector machines [17], envelope spectrum and SVM [18], time-domain features and neural networks [19], fuzzy logic [20], Statistical index development from time domain [21], fatigue life in non-stationary conditions [22] and wavelet analysis and envelope detections [23]. To provide decisions based on fault diagnosis, there are works that has implemented computing with condition monitoring techniques [24, 25].

Due to the advancements in the computing fields, there exist innumerable procedures to provide decisions based on the input and environment parameters. One of the emerging areas is the context aware systems that come under pervasive and ubiquitous computing [26]. This technology is prominent in the areas of information technology [27], mobile services [28], web services [29], internet of things [30] etc. The context-aware systems can be adapted to the existing and future possible environments without the interactions of users with effective decision making capabilities. Earlier, these context-aware systems for diagnosis have been used in Quality of Service (QoS) management [31], early diagnosis of bipolar disorders [32] and heart diseases [33]. This paper utilizes the conceptual methodology of context-aware systems for bearings in roller to provide decisions for maintenance actions by perceiving the context. There are several existed contextual models; popular is ontology based models to define the relations among the input, environment and output maintenance actions. The several input variables are programmed by using Physical model of the bearing to achieve different combinations of output variables and patterns. By perceiving these patterns, appropriate maintenance decisions can be taken on the rollers to improve performance.

2 Rolling Element Bearing and Roller in Paper Industry

The present work is carried out in BillerudKorsnäs production unit in Karlsborg, Sweden, and focus on one roller located in the wire section. There are several rollers operating in this industry out of which this work focuses on three rollers that requires main maintenance actions to be followed to increase performance. The three rollers are modeled using Physical Model in Fig. 1 using NX 8.5 tool.

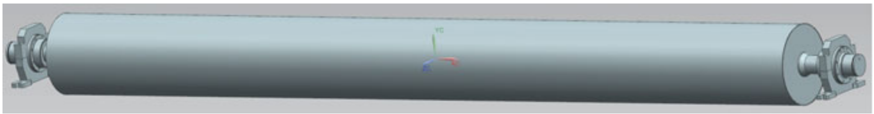


Fig. 1 Physical model of a roller in a paper mill

3 Diagnosis

IN general, diagnostics investigates or analyzes the cause or nature of condition, situation, or problem, whereas prognostics concerned with calculating or predicting the future as a result of rational study and the analysis of available pertinent data. In terms of the relationship between prognostics and diagnostics, diagnostics is the process of detecting and identifying a failure mode within a system or sub-system.

Machine fault diagnostics is a procedure of mapping the information obtained in the measurement space and/or features in the feature space to machine faults in the fault space. This mapping process is also called pattern recognition. Traditionally, pattern recognition has been done manually with auxiliary graphical tools such as power spectrum graph, phase spectrum graph, cepstrum graph, AR spectrum graph, spectrogram, wavelet scalogram, and wavelet phase graph to name a few. However, manual pattern recognition requires expertise in the specific area of diagnostic application, highly trained and skilled personnel are needed. Therefore, automatic pattern recognition is highly desirable. This can be achieved by classification of signals on the information and/or features extracted from the signals [34, 35].

Modal analysis is typically performed to determine a structure's vibration characteristics such as natural frequencies, mode shapes and mode participation factors. From the modal analysis the mode shapes and the eigen-frequencies were calculated. In this study, we performed the modal analysis on one of the rollers using FEM (Finite Element Method) for dynamic response. The result from this FEM analysis was performed in ANSYS found the 2nd bending mode that depicted in Fig. 2. Similarly, we modeled and analyzed three rollers with felt wire by ANSYS as shown in Fig. 3. The three rollers are running by the felt wire that acts as transfer mechanism of paper. This physical model was being updated dynamically by making use of continuous online condition monitoring data from vibration of the rollers in the paper mill industry. The several input parameters and output parameters were gathered simultaneously to obtain the context for the basis of contextual modeling.

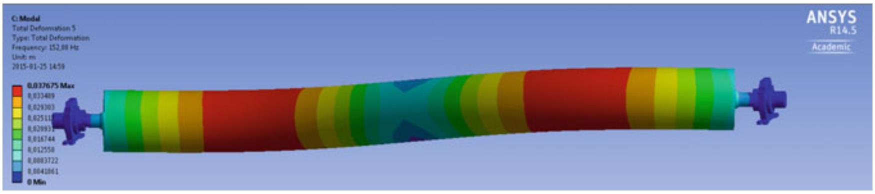


Fig. 2 Dynamic response with deformations

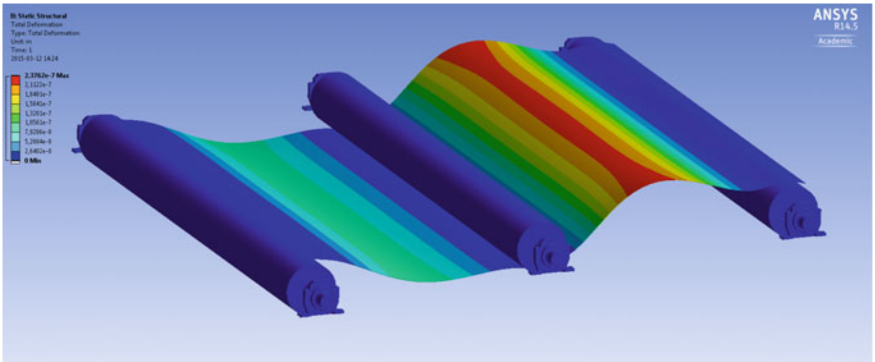


Fig. 3 Dynamic response of three rollers with felt wire

4 Ontology Based Contextual Modeling

The concept of context-aware computing is described by [27, 36] as "*the ability of a mobile user's applications to discover and react to changes in the environment they are situation*". The popular definition for the context was defined by [37] as "*any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves*". Another way to classify the context is to consider the dimensional aspects dividing context into three categories; computing context (like network connectivity, communication instances and peers), user context (like profile of user, location, people nearby and social situations) and physical context (like environment, physical devices) [27]. There is a need of model that abstracts and stores the data in meta-database for decision making for context-aware systems. One of the popular contextual models is the Ontology based model. This models use ontologies to represent the context means relationships, concepts among entities in a structure or shared domain knowledge [38]. They are becoming powerful by the applicability of formal expressiveness and reasoning techniques. In general, ontology and logic based models are not used in Wireless Sensor Networks (WSN) because of resource constraints. Ontology based models are most useful for determining relationships, dependencies and reasoning among the variables and permits little bit of heterogeneity and efficient contextual provisioning [39].

Semantic inference is the method for implementing the process of ontology for the knowledge base [40] and this inference acts as grammar of the standard form of ontology languages such as Ontology Inference Layer (OIL), Resource Description Framework Schema (RDFS) and Web Ontology Language (OWL) [41]. An ontology-based inference engine accomplishes information retrieval and question and answer (Q&A) functions by getting information about a specific instance [42].

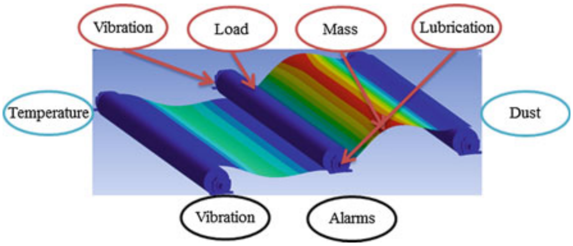


Fig. 4 Primary sensor interpretation of input, environmental and outout parameters

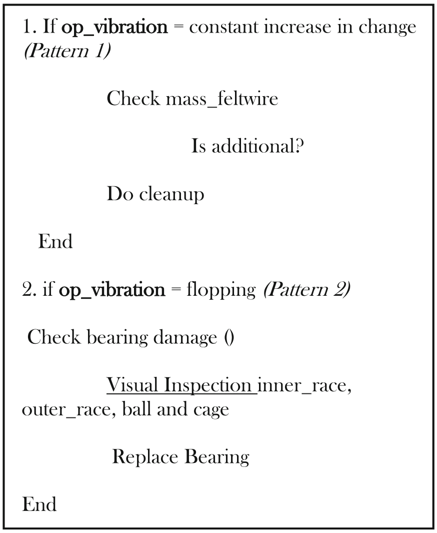
In some cases in maintenance, where i statements or procedure are included in data, there is need to process the text by using natural language process [43].

Context information is the technology of information analyzing and characterizing the real context in virtual space by involving relation between real and virtual parameters to deliver a personalized service [44]. The context sensors perceive the context using real sensors such as temperature, vibration, humidity etc. The virtual sensors interpret the factors from the context of the physical sensors. In this paper, rollers in there paper mill, there exist several SKF primary sensors acquire real information such as vibration, load, mass of dust on the felt wire, lubrication quantity on rolling element bearing, temperature, humidity and dust density as shown in Fig. 4.

The secondary sensors in this case are the rotational speeds that produce the peak amplitudes in the resonant frequency found in vibration frequency analysis. The patterns of frequency response are captured for the purpose of specific maintenance action. For example, a change in the residual mass that forms on the felt wire is replicated on the frequency response of vibration signal. If the pattern of the output response is similar to the pattern of response to high mass forming on the feltwire, the wire requires cleaning. The output alarms produce signals if there is a concern about safety, illumination, acoustic detection and leakage or breakage of the subcomponents in system. Other maintenance actions include visual inspection, repair/replacement of the items in setup, ventilation and noise cancellation depends on the output variables. An additional step is required when there is a detection of peaks from the vibration signals to acquire necessary event. These events need further investigation whether there are any failures or not. These contextual parameters are converted to the OWL for inference of rules generation.

The inference rules for service are applied to inference information about services using external context information deinferred from the External Context Ontology of maintenance [45] actions and internal context information from the Internal Context Ontology of the condition monitoring [42]. These inference rules are generated either by the previous history of maintenance actions or by input actions based on experience using contextual parameters. Some sample diagnostic rules are shown in Fig. 5.

Fig. 5 Inference rules for diagnosis based on dynamic response of rolling element bearing



```
1. If op_vibration = constant increase in change
(Pattern 1)

            Check mass_feltwire

                        Is additional?

            Do cleanup

  End

2. if op_vibration = flopping (Pattern 2)

 Check bearing damage ()

            Visual Inspection inner_race,
outer_race, ball and cage

             Replace Bearing

End
```

5 Architecture of Ontology Driven Diagnosis

The configuration of ontology based diagnosis for maintenance decisions (ObDMD) architecture that is modified from [42] is shown in Fig. 6. The data acquisition layer acquires data from condition monitoring data and sensors for primary data. This vibration data from physical model is then converted to OWL. If there is an event occurs, the data from the sensors are acquired. External context information can also be accessed from data storage layer. The context query is then transferred to context layer based on the event generation and thus create context instance. These instances are converted to ontology models in data storage layer.

In the inference layer, the created context instances are provided to the context manager that creates relations among the maintenance actions mapped by context mapping from the knowledge layer as shown in Fig. 5. In Service Mapping, for determining the format of the inferred results of service, data is converted into a format required in each application through the predetermined service content database. This rules are created in RDF format. The reasoner decides the reasoning of each of the rules based on weights of maximum impact of each action. In the event of a conflict of rules, the reasoner provides the best maintenance output action that increases the performance using inputs from the data storage and knowledge layer. Various computing techniques such as neural networks, decision trees, fuzzy rules and statistical techniques can be used to provide optimum action. The service layer thus provides access to the application layer for an interactive informatiton on diagnosis of the roller using context-awareness. Each of the knowledge and data storage layers is updated regularly and triggered in the triggering of an event. Out of

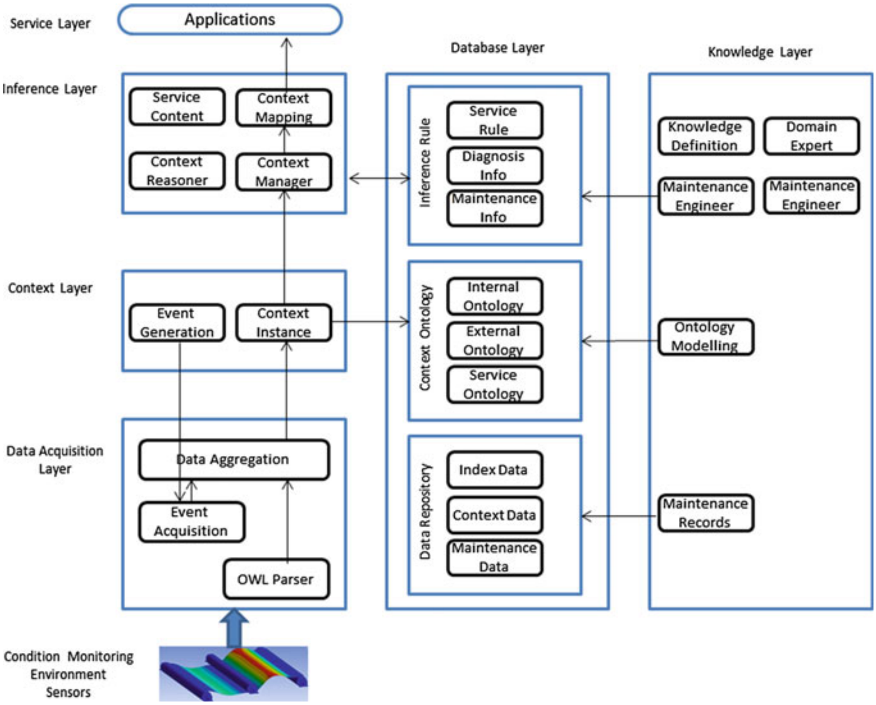


Fig. 6 Ontology based diagnosis for maintenance decision (ObDMD) architecture

all the layers, the inference layer guides maintenance recommendation using the information from all the other layers and it suggests the necessary actions to the user.

6 Conclusions

This paper proposes the conceptual application of a context-aware decision model to a paper mill roller using the dynamic response obtained from a physical model to build the ontology model. The proposed process represents in using computing techniques for the purpose of maintenance diagnosis in the paper industry. To fully implement this mechanism, we require rules from the users in the paper industry, research experts, we also need to know the history of failures and maintenance actions and have condition monitoring data to auto detect the anomalies in the frequency response. If necessary maintenance actions are taken at the appropriate time, operating costs will decrease and performance and production will increase. If we make use of existing data inferences and combine these with modelling the remaining useful life, we may able to perform prognosis by combing with modelling of remaining useful life.

References

1. Zhang XN (1998) Research on the condition monitoring and forecasting for large rotating machine. Ph.D. thesis, Department of Mechanical Engineer, Xi'an Jiaotong University, Xi'an, China
2. Winder RL, Littmann WE (eds) (1976) Bearing damage analysis. National Bureau of Standard Publication, Washington, D.C
3. He ZJ, Zi YY, Meng QF (2001) Fault diagnosis principle of non-stationary signal and applications to mechanical equipment. Higher Education Press, Beijing
4. Qu LS, He ZJ (1986) Mechanical fault diagnostics. Shanghai Science and Technology Press, Shanghai
5. Tandon N, Choudhury A (1999) A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. Tribol Int 32:469–480
6. Sugumaran V, Ramachandran KI (2007) Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing. Mech Syst Signal Process 21(5):2237–2247
7. Yam RCM, Tse PW, Li L, Tu P (2001) Intelligent predictive decision support system for condition-based maintenance. Int J Adv Manuf Technol 17(5):383–391
8. Al-Najjar B, Wang W (2001) A conceptual model for fault detection and decision making for rolling element bearings in paper mills. J Qual Maintenance Eng 7(3):192–206
9. 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
10. Al-Najjar B, Alsyouf I (2003) Selecting the most efficient maintenance approach using fuzzy multiple criteria decision making. Int J Prod Econ 84(1):85–100
11. Rafsanjani A et al (2009) Nonlinear dynamic modeling of surface defects in rolling element bearing systems. J Sound Vib 319(3):1150–1174
12. While MF (1979) Rolling element bearing vibration transfer characteristics: effect of stiffness. J Appl Mech 46(3):677–684
13. Liew A, Feng NS, Hahn EJ (2001) Transient rotordynamic modeling of rolling element bearing systems. In: ASME Turbo Expo 2001: power for land, sea, and air. American Society of Mechanical Engineers
14. Gryllias KC, Antoniadis IA (2012) A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments. Eng Appl Artif Intell 25(2):326–344
15. Nikolaou NG, Antoniadis IA (2002) Rolling element bearing fault diagnosis using wavelet packets. NDT E Int 35(3):197–205
16. Gebraeel N, Lawley M, Liu R, Parmeshwaran V (2004) Residual life predictions from vibration-based degradation signals: a neural network approach. IEEE Trans Ind Electron 51(3):694–700
17. Yang J, Zhang Y, Zhu Y (2007) Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension. Mech Syst Signal Process 21(5):2012–2024
18. Yang Y, Yu D, Cheng J (2007) A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. Measurement 40(9):943–950
19. Sreejith B, Verma AK, Srividya A (2008) Fault diagnosis of rolling element bearing using time-domain features and neural networks. In: IEEE Region 10 and the third international conference on industrial and information systems, 2008, ICIIS 2008. IEEE, pp 1–6
20. Liu TI, Singonahalli JH, Iyer NR (1996) Detection of roller bearing defects using expert system and fuzzy logic. Mech Syst Signal Process 10(5):595–614
21. Fuqing Y, Kumar U (2014) Statistical index development from time domain for rolling element bearings. Int J Perform Eng 10(3):313
22. Leturiondo U, Salgado O, Galar D, Mishra M (2014) Methodology for the estimation of the fatigue life of rolling element bearings in non-stationary conditions. In: International conference on condition monitoring of machinery in non-stationary operations, Lyon, France, 15–16 Dec 2014

23. Peter WT, Peng YH, Yam R (2001) Wavelet analysis and envelope detection for rolling element bearing fault diagnosis—their effectiveness and flexibilities. J Vib Acoust 123(3):303–310
24. Cococcioni M, Lazzerini B, Volpi SL (2009) Rolling element bearing fault classification using soft computing techniques. In: IEEE international conference on systems, man and cybernetics, 2009, SMC 2009. IEEE, pp 4926–4931
25. Gs V, Sriram NS (2010) A comparative study of soft computing techniques for rolling element bearing condition monitoring using vibration signal analysis. In: International conference ICETMCA 2010 (International Conference on Emerging Trends in Mathematics and Computer Applications)
26. Weiser M (1993) Some computer science issues in ubiquitous computing. Commun ACM 36(7):75–84
27. Schilit B, Adams N, Want R (1994) Context-aware computing applications. In: First workshop on mobile computing systems and applications, 1994, WMCSA 1994. IEEE, pp 85–90
28. Chen G, Kotz D (2000) A survey of context-aware mobile computing research, vol 1, no 2.1. Technical report TR2000-381, Department of Computer Science, Dartmouth College, 2000
29. Truong H-L, Dustdar S (2009) A survey on context-aware web service systems. Int J Web Inf Syst 5(1):5–31
30. Perera C, Zaslavsky A, Christen P, Georgakopoulos D (2014) Context aware computing for the internet of things: a survey. Commun Surv Tutorials IEEE 16(1):414–454
31. Lin X, Cheng B, Chen J (2010) Context-aware end-to-end QoS qualitative diagnosis and quantitative guarantee based on Bayesian network. Comput Commun 33(17):2132–2144
32. Tacconi D et al (2007) On the feasibility of using activity recognition and context aware interaction to support early diagnosis of bipolar disorder. In: Proceedings of Ubiwell 2007: Ubicomp Workshop Proceedings
33. Kim TS, Kim H-D (2005) Context-Aware computing based adaptable heart diseases diagnosis algorithm. In: Knowledge-based intelligent information and engineering systems. Springer, Berlin
34. Mishra M, Saari J, Galar D, Leturiondo U (2014) Hybrid models for rotating machinery diagnosis and prognosis: estimation of remaining useful life. Luleå: Luleå tekniska universitet. (Technical report, Luleå University of Technology)
35. Korbicz, J (ed) (2003) Fault diagnosis: models, artificial intelligence, applications. Springer, Berlin
36. Schilit BN, Theimer MM (1994) Disseminating active map information to mobile hosts. Netw IEEE 8(5):22–32
37. Dey AK, Abowd GD, Salber D (2001) A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. Hum Comput Inter 16(2):97–166
38. Öztürk P, Aamodt A (1997) Towards a model of context for case-based diagnostic problem solving. In: Context-97; Proceedings of the interdisciplinary conference on modeling and using context, 1997
39. Thaduri A, Kumar U, Verma AK (2015) Computational intelligence framework for context-aware decision making. Int J Syst Assur Eng Manage 6:1–12
40. Russell S, Norvig P (1995) Artificial intelligence: a modern approach. Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs 25 (1995)
41. Cimiano P (2006) Ontology learning from text. Springer, US
42. Kim J, Kim J, Lee D, Chung KY (2014) Ontology driven interactive healthcare with wearable sensors. Multimedia Tools Appl 71(2):827–841
43. Busch JE, Lin AD, Graydon PJ, Caudill M (2006) Ontology-based parser for natural language processing. US Patent 7,027,974 11 Apr 2006
44. Jung-Hyun LEE (2004) User preference mining through hybrid collaborative filtering and content-based filtering in recommendation system. IEICE Trans Inf Syst 87(12):2781–2790
45. Galar D, Thaduri A, Catelani M, Ciani L (2015) Context awareness for maintenance decision making: a diagnosis and prognosis approach. Measurement 67:137–150

Context Awareness in Predictive Maintenance

Bernard Schmidt, Diego Galar and Lihui Wang

Abstract Maintenance of assembly and manufacturing equipment is crucial to ensure productivity, product quality, on-time delivery, and a safe working environment. Predictive Maintenance approach utilizes the condition monitoring (CM) data to predict the future machine conditions and makes decisions upon this prediction. Recent development in CM leads to context aware approach where in parallel with CM measurements also data and information related to the context are gathered. Context could be operational condition, history of machine usage and performed maintenance actions. In general more obtained information gives better accuracy of prediction. It is important to track operational context in dynamically changing environment. Today in manufacturing we can observe shift from mass production to mass customisation. This leads to changes from long series of identical products to short series of different variants. Therefore implies changing operational conditions for manufacturing equipment. Moreover, where asset consist of multiple identical or similar equipment the context aware method can be used to combine in reliable way information. This should allow to increase accuracy of prediction for population as a whole as well as for each equipment instances. Same of those data have been already recorded and stored in industrial IT systems. However, it is distributed over different IT systems that are used by different functional units (e.g. maintenance department, production department, quality department, tooling department etc.). This paper is a conceptual paper based on

B. Schmidt (✉) · D. Galar · L. Wang
University of Skövde, PO Box 408, 541 28 Skövde, Sweden
e-mail: bernard.schmidt@his.se

D. Galar
e-mail: diego.galar@ltu.se

L. Wang
e-mail: lihuiw@kth.se

D. Galar
Luleå University of Technology, 971 87 Luleå, Sweden

L. Wang
KTH Royal Institute of Technology, 100 44 Stockholm, Sweden

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_15

initial research work and investigation in two manufacturing companies from automotive industry.

Keywords Context modeling · Context awareness · Condition monitoring · Condition based maintenance · Predictive maintenance

1 Introduction

Maintenance is crucial to ensure reliability of assembly and manufacturing equipment thereafter productivity, product quality, on-time delivery, and a safe working environment. Implementation of effective prognosis for maintenance can bring variety of benefits including increased system safety, improved operational reliability, increased maintenance effectiveness, reduced maintenance inspection and repair-induced failure, and reduced lifecycle cost [1].

Maintenance approaches in industrial history evolve [2] and it is an ongoing process. At earlier stages the Corrective Maintenance also known as reactive maintenance or run-to-failure was used. Later approach called Preventive Maintenance (PM) was focused on taking actions before the failure occurs. This approach evolved to Condition Based Maintenance (CBM), where the decisions are made based on the machine condition indicators obtained in most cases through measurement systems. Predictive Maintenance (PdM) and Prognostics and Health Management (PHM) are approaches that utilize the condition monitoring data to predict the future machine health state and make decisions upon this prediction.

Nowadays in quickly developing word we are facing new challenges and opportunities.

The paradigm of mass customization aims to deliver customized products with near mass production efficiency. Mass customization is imperative for many companies to survive in the fragmented, diversified, and competitive marketplace [3]. Frequent changes in produced variants imply changes in operational conditions of manufacturing equipment.

Internet of Things (IoT) is a paradigm where everyday objects are connected to the Internet. It allows devices communication with each other with minimum human intervention [4]. The term has been initially used by Kevin Ashton in 1999. In [5] he describes the IoT as an enabler to know when things need replacing, repairing or recalling.

However, large number of smart devices and sensors is producing huge amount of data that need to be processed in a useful way, as data is not useful unless it is processed in a way that provides context and meaning that can be understood by the right personnel [6]. Those aggregated streams of data, are called “Big Data”.

Cloud Manufacturing (CMfg) paradigm is a result of combination of cloud computing, the Internet of things, service-oriented technologies and high performance computing [7]. It transforms manufacturing resources and capabilities into manufacturing services. It is not simple deployment of manufacturing software tools in the computing cloud. The physical resources integrated in the manufacturing cloud are able to offer adaptive, secure and on-demand manufacturing services over the Internet of Thinks [8]. Effect of this paradigm on the maintenance approach is not well elaborated in the literature. Nevertheless we can imply that delivering manufacturing resources as a service may cause more dynamic changes in operational conditions of manufacturing assets.

Recently context awareness approach is gaining focus of researchers from the field of CBM and PdM. This well-known concept in some other fields, could be beneficial when employed in CBM/PdM.

The rest of the paper is organized as follows. Section 2 provides the details about context and its different modeling techniques and overview on the concepts of context-aware systems; Sect. 4 depicts the context-awareness in the context of Predictive Maintenance; and finally, Sect. 4 discusses and concludes the paper.

2 Context

The popular definition of the context according to [9] was defined by [10] as “any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves”.

Noticeable amount of research on context in relation to context-aware systems comes from pervasive and mobile computing research area. This could be because of the popularity that mobile devices and mobile applications have gained in recent years.

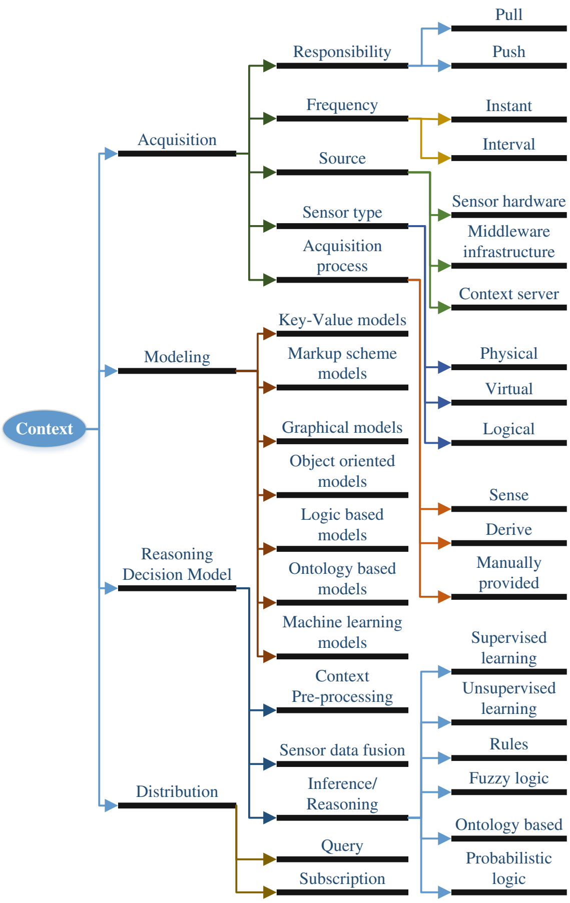
As Artificial Intelligence (AI) methods are also option for Predictive maintenance [11] it could be fruitful to check what are the contribution from that field. In AI study of a formal notation of context has a long history. In depth comparison of two main formalizations from a technical and a conceptual point of view have been presented by Serafini and Bouguet in [12]. Overview of those formal definitions of context from Artificial Intelligent field are presented in Sect. 2.1.5.

Context related aspects based on [4, 13] has been presented in Fig. 1. Some of them has been elaborate more in following sections and subsections.

2.1 Context Modeling

Several context modeling technique are used in context-aware computing [13–15]. Each of those techniques has strengths and weaknesses, so incorporating multiple modeling techniques brings efficient and effective results [4].

Fig. 1 Aspects of context aware systems



2.1.1 Key-Value Models

Key Value Models (KVM) uses 2-tuple data structure <key,value>, that represents identifier (name) of the attribute and its value. Among all other implementations it is the simplest form of context representation. With small amount of data they are easy to manage. However, key-value modeling is not suitable to represent hierarchical structures or relationships.

2.1.2 Markup Scheme Models

Markup Scheme Models (MSM) is an extension over KVM. It use markup tags and hierarchical data structure. One of the advantages of this modeling is efficient data retrieval. It also support validation through schema definitions. Popular technique for markup schemas is XML [4]. It is widely used to store temporary data and transfer data among applications or application components in. Markup schema modeling can be performed in any language or mechanism that supports tag based storage e.g. JSON [16].

2.1.3 Graphical Models

Graphical Models (GM) represents context with relationships. Modeling techniques that can be employed are Unified Modeling Language (UML) or Object Role Modeling (ORM). The advantage of GM over KVM and MSM is that it allows to capture relationships in the context model. Examples of low-level representations of graphical modeling could be a SQL database, noSQL database, XML.

2.1.4 Object Oriented Models

Object Oriented Models (OOM) uses class hierarchies and relationships. It can be easily integrated into context-aware systems, as most of the high-level programing languages support object oriented concepts. However, due to the lack of standards, the validation of object oriented design is difficult. Moreover it does not provide inbuilt reasoning capabilities.

2.1.5 Logic Based Models

In Logic Based Models (LBM) context is represented with use of facts, expressions and rules. It allows creation of new high-level context using low-level context.

Propositional Logic of Context (PLC)

- Contexts are first class objects. The formal language of a theory of context should contain terms denoting contexts, and should allow one to predicate properties about these objects and to express relations between contexts (e.g., that one context is more general than another), or between contexts and other objects (e.g., that the time of a context c is t).
- A formula is always stated in a context. However, the same context can be described from different perspectives, i.e., the content of a particular context is itself context dependent. This property is called non-flatness, and each formula have to be prefixed by a sequence of context labels e.g. $\kappa_1 \ldots \kappa_n$: φ.
- A context is modeled as a set of truth assignments, each of which represents a possible state of the world as described in the context.
- A context is always partial. Only a subset of what can be said is given an interpretation in each context.

- Statements about a context are stated in other contexts via so-called ist-formulas, i.e., formulas of the form ist(κ,φ). The formula ist(κ,φ) is read as "φ is true in the context κ". This formula, if asserted in a context κ', means that, viewed from κ', φ is true in κ.
- There is an intuitive relation between the assertions κ'κ:φ and κ':ist(κ,φ). Indeed, the latter is true if the former is true, and vice versa. This property is axiomatized via an inference rule called CS (a contextual version of the modal rule of necessitation) that allows deriving κ':ist(κ,φ) from κ'κ:φ. This is the main contextual reasoning pattern allowed in PLC.
- Other relations between contexts can be stated through lifting axioms which relate the truth in one context to the truth in another context.
- Like any other formula, lifting axioms are always stated in a context, called an outer context.
- There is no outermost context. For any context κ, there is an outer context κ'from which κ can be described.

Local Model Semantic/MultiContext Systems (LMS/MCS)

- A context is primarily a subset of a partial and approximate theory of the world from some individual's perspective. The collection of facts used to reason about a given problem by individuals is the most typical example.
- Reasoning mainly happens locally to a single context. Only those facts relevant to the problem individuals want to solve are taken into consideration.
- There are possible relations between local reasoning processes, as different contexts are not simply unrelated representations, but different representations of the same world. For example, two contexts may describe the same piece of world at different level of detail from the same perspective; or may describe it, only from different perspectives. Relations between different perspectives in LMS are represented via a compatibility relation between local interpretations associated with each context. The proof theoretic counterpart of compatibility relations are bridge rules, i.e., inference rules with premises and consequences in different contexts.
- The relationship between different contexts, in general, can be described only to a partial extent, as each of them may encode assumptions which are not fully explicit.

Comparison

The main feature of a formal theory of the context, is the ability to formalize the relations existing between different contexts. To achieve this goal, PLC and LMS/MCS adopt two different strategies:

- PLC is based on a combination of lifting axioms as well as axioms and rules for exiting and entering contexts.
- LMS/MCS is based on the mechanism of bridge rules.

Example, how PLC and MCS/LMS represent the fact that ψ in κ is a logical consequence of φ in κ' has been showed in Fig. 2. In PLC, one needs a third ("top") context κ" where logical consequence is represented by the formula ist$(\kappa,\varphi) \supset$ ist (κ',ψ) Fig. 2a. Instead, in MCS this is directly represented by the fact that $\kappa' : \psi$ is derivable via bridge rules from the assumption $\kappa{:}\varphi$, i.e., that $\kappa{:}\varphi \models_{MCS} \kappa'{:}\psi$ Fig. 2b.

2.1.6 Ontology Based Models

Term ontology comes from philosophy where it refers to a theory of the nature of existence. In computer and information science, ontology determines formal specifications of knowledge in a domain explicit specification of the objects, concepts, and other entities (vocabulary) that exist in some area of interest ant the relationships that hold among them [17].

According to [18] ontology based context modeling allows: ❶ knowledge sharing between computational entities by having common set of concepts about concept; ❷ logic inference by exploiting various existing logic reasoning mechanisms to deduce high-level, conceptual context from low-level, raw context; ❸ knowledge reuse by reusing well-defined Web ontologies of different domains, e.g. large-scale context ontology can be composed without starting from scratch.

Web Ontology Language (OWL) is modeled through an object-oriented approach, where structure of a domain is described in terms of classes and properties.

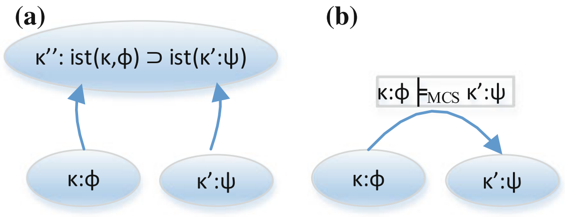


Fig. 2 Inference in PLC and LMS/MCS

2.1.7 Machine Learning Models

Machine Learning Models (MLM) use machine learning techniques. It is not a strictly context modeling approach, however does target similar objectives. It has been presented in [19] to enable effective personalized service provision.

MLM has been indicated in [9] as the best approach for intelligent context-aware system.

2.2 Context Reasoning

Context reasoning, also called inference, can be defined as method of deducing new knowledge, that can be also understand as high-level context, based on the available context [4]. Context reasoning techniques could be classified into six categories as in Fig. 1: supervised learning (e.g. Artificial Neural Network, Bayesian Networks, Case-based Reasoning, Decision Tree Learning, Support Vector Machines), unsupervised learning (e.g. Clustering, k-nearest Neighbour), rules, fuzzy logic, ontology based, and probabilistic logic (e.g. Dempster-Shafer, hidden Markov Models, naïve Bayes). There is relationship between context reasoning and context modeling as some reasoning techniques prefer some modeling techniques [4]. Imperfection and uncertainty of a raw context are the factors that also emerged the requirement of reasoning step. Fuzzy Logic and Probabilistic logic has been indicated as reasoning techniques that can handle uncertainty.

It has been revealed in [14] that different models and techniques needs to be integrated with each other within hybrid context modeling approach in order to obtain more general and flexible systems.

Perera et al. in [4] provided example of the hybrid context modeling and reasoning approach. Statistical techniques can be used on lowest level to fuse sensor data. Further, fuzzy logic could be used to convert fixed data into more natural terms. Dempster-Shafer can be used to combine sensor data from different sources. Machine learning techniques such as artificial neural networks and support vector machines can be used for further reasoning. Thereafter, the high level data can be modeled using semantic technologies as ontologies.

2.3 Context-Aware Systems

The context-aware systems can be defined as systems that are adaptable to the existing and future possible environments without the interactions of users [9]. In the pervasive computing community there is a growing body of research on the use of context-awareness as a technique for developing pervasive computing applications that are flexible, adaptable, and capable of acting autonomously on behalf of users [14].

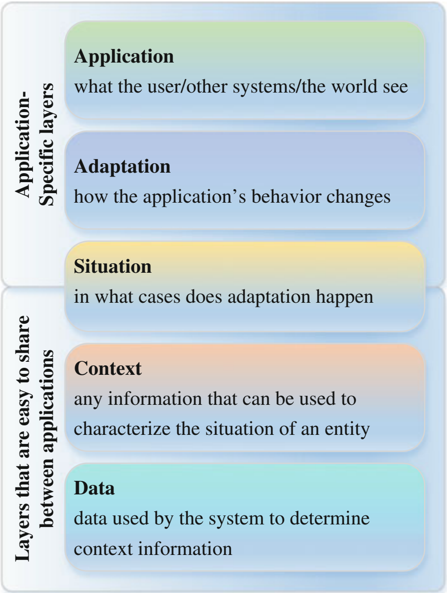

Application-
Specific layers
Layers that are easy to share
between applications
Application
what the user/other systems/the world see
Adaptation
how the application's behavior changes
Situation
in what cases does adaptation happen
Context
any information that can be used to
characterize the situation of an entity
Data
data used by the system to determine
context information

Fig. 3 Anatomy of context-aware application based on [20]

Context-aware systems most often are represented with use of layered architecture. Context models can be seen as an abstraction layer between applications and the technical infrastructure that provides the context data [21]. In [22] four layers are depicted: ❶ sensor layer for data acquisition, ❷ data storage layer for maintaining data, ❸ processing layer for all analysis and modeling, and ❹ application layer for final representation.

A little bit different structure has been proposed in [20] that is presented in Fig. 3. In this approach layers has been grouped into application specific layers and layers that can be shared among different applications.

Architecture structure presented in [9] consists of interacting building blocks as: ❶ data acquisition layer, ❷ pre-processing layer, ❸ network layer, ❹ data storage layer, ❺ decision and control layer, and ❻ user interface layer.

3 Context in PdM

Concepts of "context" and "context aware system" have been not well utilized by researchers from field of CBM and PdM. Analysis of ten review and survey papers in the area from time period 2005–2014 mentioned in literature review [23] reveals that term "context" in the context of predictive maintenance is never directly mentioned. However, some of indicated challenges and future trends can be addressed by utilization of "context" and "context-awareness" approach. Some of those are: more basic and applied research in decision making systems [24]; data fusion of multi-dimensional CM, model the influence of external environmental variables, deal with multiple failure modes [25]; needed of general methodology for prognostics [11]; the development of methods or tools for extraction, processing and interpretation of knowledge type information [26]; the consideration of the effects from maintenance actions, the consideration of failure interactions [27].

Integration of disparate data sources that are commonly available in industry has been proposed for better maintenance decision making [28]. The cloud approach is pointed as a feasible solution for this integration, and XML language is presented as a tool that can be used for data integration.

Lee et al. [6] indicated that algorithms can perform more accurately when more information throughout the machine's lifecycle, such as system configuration, physical knowledge and working principles, are included, so there is a need to systematically integrate, manage and analyse machinery or process data during different stages of machine life cycle.

Recently, there are reported works that apply context approach in the field of maintenance.

In [29] the context-aware approach has been used in energy domain for predicting the future load. The prediction model parameters are stored in repository with context in which they were valid this allow to retrieve them when similar context occurs. Repository has been organized as binary search tree.

Semantic and modeling for a contextualized mobile client of a distributed model that constitutes a maintenance mobile cloud has been presented Pistofidis and Emmanouilidis [30]. Presented WelCOM platform utilizes smart sensor infrastructure for machine condition monitoring and to deliver a context aware asset management tool has been interfaced with Computerized Maintenance Management System. Authors pointed that context modeling, identification and context-based adaptation are key elements in WelCOM approach. Five context categories has been identified for mobile maintenance advisor: ❶ user context—information about role, expertise, activity, location, preferences; ❷ system context—device specification, network status, security profile, energy consumption; ❸ environment context—sensor readings (temperature, noise, light), user proximity, asset proximity, timestamp; ❹ service context—criticality, priority, task sequence, dependencies, constrains, support; ❺ social context—group/team participation, relationship role, interaction profile, rank.

In fleet-based approach presented in [31], an ontology model has been proposed with a following context types: ❶ technical context—technical features and characteristics of the system/sub-system/equipment; ❷ dysfunctional context—the generic degradation modes on the units; ❸ operational context—operational conditions that are given by the mission to be performed for units as well as the environment that surround them; ❹ service context—usage of the unit; and ❺ application context—for maintenance optimization, enables data/model retrieval of the monitored units with its corresponding context. Than comparison of heterogeneous units could be performed based on similarity of the context. This enables data capitalization that could improve prognostics model and precision. This approach has been applied in naval domain to fleet of ships.

In [32] the context driven remaining useful life estimation has been presented. Health condition of machine is represented by so called fingerprint, while context is represented by monitored operational data that describes the way the machine has been used.

Galar et al. [22] proposed a hybrid model-based maintenance decision system with consideration of context-driven aspects. The system aims to integrate expert knowledge, physics of failure models and data driven models.

Context awareness seems to be also important from perspective of another recent trend that is application of the Cloud concept.

3.1 Cloud-Based Approach

Recently the concept of Cloud gains on popularity in research community and there has been a trend of applying cloud computing model in manufacturing industry [33].

Bahga and Madisetti presented in [34] usage of this concept in maintenance. They claimed that this is the first reported usage of the cloud architecture for maintenance data storage, processing and analysis. Their proposed hybrid approach uses local nodes for real-time fault prediction and a cloud for massive data organization and analysis.

Lee et al. [35] presented methodology and framework for a cloud-based prognostics and health management system for manufacturing industry. The system utilizes modularized algorithms as basic components to form different workflows. Workflows for typical components and mechanical problems are saved in a knowledge base that can be later used as templates for similar problems. Based on specific need (e.g. type of component for monitoring, type of data available, etc.) certain workflow can be selected and provisioned into a virtual machine as an individual Prognostics and Health Management server dedicated to an industrial user. Summarizing, in this concept the application is adapted to the specific need of the user.

4 Discussion and Conclusions

One of the issue in application of Predictive Maintenance in manufacturing industry are so called Islands of knowledge [36] that could be treated as a different contexts according to formal definitions described in chapter 2.1.5. In big manufacturing companies often there are dedicated departments focused on different aspects as production, quality, tooling, lubrication, maintenance. Those are different contexts with own specific vocabularies, reasoning and assumptions. However, as are concerned about the same piece of Word, the production line, they are not completely independent. It could be possible to identify some compatibility relations between those contexts. This could provide better, more comprehensive view that can be used for improvements in applied models and techniques in each of those areas so in Predictive Maintenance as well. The work that need to be done is to find those correlations as well as find the way to obtain those information from disparate data sources [22].

Dealing with large amount of information from disparate data sources is in concern of Big Data management. In [37] issues like how to store, integrate and

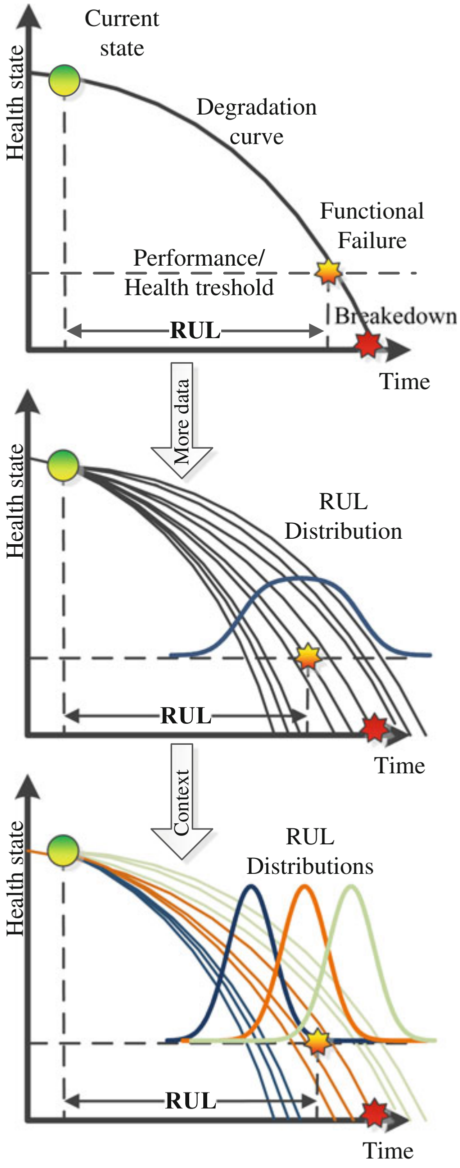


Fig. 4 Possible context-driven improvements in estimation of remaining useful life (RUL)

process those data, and how to do this in an effective and efficient way have been pointed out. Advances in this area will support context-aware approach as well.

At the end we want to summarize potential benefits that can be seen in the context driven approach and in adaptation of outcomes from other research fields that has longer research history in the area of context-awareness:

- improvement in knowledge management for later reuse,
- capitalization of data in fleet wide approach to increase accuracy and performance of algorithms,
- enabled adaptation to user needs, required in cloud-based approach,
- improved automatic selection of proper approach (e.g. signal selection, processing algorithms etc.).

In Fig. 4 we present hypothetical case how applying context-awareness can improve estimation of Remaining Useful Life (RUL) in Predictive Maintenance.

To conclude, in this paper we provide overview of the concept of context from different research field. Need for context aware application for Predictive Maintenance has been indicated, as well as recent research that utilize the context concept in it.

References

1. Bo S, Shengkui Z, Rui K, Pecht MG (2012) Benefits and challenges of system prognostics. IEEE Trans Reliab 61(2):323–335
2. Alsyouf I (2007) The role of maintenance in improving companies' productivity and profitability. Int J Prod Econ 105(1):70–78
3. Tseng M, Hu SJ (2014) Mass Customization. In: Laperrière L, Reinhart G (eds) CIRP encyclopedia of production engineering. Springer, Berlin pp 836–843
4. Perera C, Zaslavsky A, Christen P, Georgakopoulos D (2014) Context aware computing for the internet of things: a survey. Commun Surv Tutorials IEEE 16(1):414–454
5. Ashton K (2009) That 'internet of things' thing. RFID J 22:97–114
6. Lee J, Lapira E, Bagheri B, Kao H-A (2013) Recent advances and trends in predictive manufacturing systems in big data environment. Manufact Lett 1(1):38–41
7. Zhang L, Luo Y, Tao F, Li BH, Ren L, Zhang X, Guo H, Cheng Y, Hu A, Liu Y (2014) Cloud manufacturing: a new manufacturing paradigm. Enterp Inf Syst 8(2):167–187
8. Wang L, Wang XV, Gao L, Váncza J (2014) A cloud-based approach for WEEE remanufacturing. CIRP Ann Manuf Technol 63(1):409–412
9. Thaduri A, Kumar U, Verma A (2014) Computational intelligence framework for context-aware decision making. Int J Syst Assur Eng Manag 1–12
10. Dey AK, Abowd GD, Salber D (2001) A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. Hum Comput Inter 16(2):97–166
11. Peng Y, Dong M, Zuo MJ (2010) Current status of machine prognostics in condition-based maintenance: a review. Int J Adv Manuf Technol 50(1–4):297–313
12. Serafini L, Bouquet P (2004) Comparing formal theories of context in AI. Artif Intell 155(1–2): 41–67

13. Moore P, Bin H, Xiaomei Z, Campbell W, Ratcliffe M (2007) A survey of context modelling for pervasive cooperative learning. In: Proceedings of information first ieee international symposium on technologies and applications in education, pp K5-1–K5-6
14. Bettini C, Brdiczka O, Henricksen K, Indulska J, Nicklas D, Ranganathan A, Riboni D (2010) A survey of context modelling and reasoning techniques. Pervasive Mob Comput 6 (2):161–180
15. Baldauf M, Dustdar S, Rosenberg F (2007) A survey on context-aware systems. Int J Ad Hoc Ubiquitous Comput 2(4):263–277
16. JSON (2015) http://json.org/. Accessed 1 Jun 2015
17. Gruber T (2009) Ontology. Liu L, Özsu MT (eds) Encyclopedia of database systems. Springer, US, pp 1963–1965
18. Wang XH, Da Qing Z, Tao G, Pung HK (2004) Ontology based context modelling and reasoning using OWL. In: Proceedings of the second ieee annual conference on pervasive computing and communications workshops, pp 18–22
19. Flanagan JA (2005) Context awareness in a mobile device: ontologies versus unsupervised/supervised learning, pp 167–170
20. Nicklas D, Koppaetzky N (2013) Data, context, situation: on the usefulness of semantic layers for designing context-aware systems. In: Cordeiro J, Hammoudi S, van Sinderen M (eds) Software and data technologies. Springer, Berlin, pp 3–18
21. Bangemann T, Rebeuf X, Reboul D, Schulze A, Szymanski J, Thomesse JP, Thron M, Zerhouni N (2006) PROTEUS-Creating distributed maintenance systems through an integration platform. Comput Ind 57(6):539–551
22. Galar D, Thaduri A, Catelani M, Ciani L (2015) Context awareness for maintenance decision making: a diagnosis and prognosis approach. Measurement 67:137–150
23. Schmidt B, Wang L (2015) Predictive maintenance: literature review and future trends. In: Proceedings of FAIM
24. Teti R, Jemielniak K, O'Donnell G, Dornfeld D (2010) Advanced monitoring of machining operations. CIRP Ann Manuf Technol 59(2):717–739
25. Si X-S, Wang W, Hu C-H, Zhou D-H (2011) Remaining useful life estimation—a review on the statistical data driven approaches. Eur J Oper Res 213(1):1–14
26. 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
27. 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
28. Galar D, Kumar U, Juuso E, Lahdelma S (2012) Fusion of maintenance and control data: a need for the process. In: Proceedings of 18th world conference on nondestructive testing
29. Dannecker L, Schulze R, Böhm M, Lehner W, Hackenbroich G (2011) Context-aware parameter estimation for forecast models in the energy domain. In: Bayard Cushing J, French J, Bowers S (eds) Scientific and statistical database management. Springer, Berlin, pp 491–508
30. Pistofidis P, Emmanouilidis C (2013) Profiling context awareness in mobile and cloud based engineering asset management. In: Emmanouilidis C, Taisch M, Kiritsis D (eds) Advances in production management systems. competitive manufacturing for innovative products and services. Springer, Berlin, pp 17–24
31. Medina-Oliva G, Voisin A, Monnin M, Kosayyer N, Léger JB (2013) A case study for fleet-wide semantic based predictive diagnostic. In: Proceedings of the European safety and reliability conference, ESREL 2013, pp 1751–1760
32. Johansson C-A, Simon V, Galar D (2014) Context driven remaining useful life estimation. Procedia CIRP 22:181–185
33. Xu X (2012) From cloud computing to cloud manufacturing. Robot Comput Integr Manuf 28 (1):75–86
34. Bahga A, Madisetti VK (2012) Analyzing massive machine maintenance data in a computing cloud. IEEE Trans Parallel Distrib Syst 23(10):1831–1843

35. Lee J, Yang S, Lapira E, Kao H-A, Yen N (2013) Methodology and framework of a cloud-based prognostics and health management system for manufacturing industry. Chem Eng Trans 33:205–210
36. Schmidt B, Wang L (2015) Cloud-based predictive maintenance. In: Proceedings of FAIM
37. Assunção MD, Calheiros RN, Bianchi S, Netto MAS, Buyya R (2015) Big data computing and clouds: trends and future directions. J Parallel Distrib Comput 79–80:3–15

Prognostics and Health Management: Methodologies & Soft Computing Techniques

S.V. Shrikhande, P.V. Varde and D. Datta

Abstract For safety systems of Indian nuclear plants, mean life estimates are found using MIL-STD-217FN2 or RAIC-HDBK-217Plus. These statistical life estimates varies from item to item due to statistical variations in base material defects, fabrication, operational stresses and use environment. For achieving the system reliability, replacements done based on these life estimates results in under utilization of its complete life. On the contrary when failure happens earlier than this estimate, it is expected that failures are identified by online self diagnostic. To reveal hidden failures which are not detected by online self diagnostics, periodic surveillance tests are done. These discovered faults need immediate repair attention which is unscheduled maintenance. Also recent computer based systems uses programmable devices like FPGA/CPLD which are identical in redundant trains and therefore susceptible to Common Cause Failures. To overcome these difficulties prognostics giving indication of the impending failure and estimate of remaining useful life is important so that planned scheduled maintenance can be carried out. Prognostics require in situ monitoring sensors, data collection, pre-processing for feature extraction, damage assessment and remaining life estimation by soft computing techniques. This paper suggests techniques of monitoring degradation for CMOS electronic components and feature extraction. This paper discusses soft computing techniques of Support Vector Machines for classification (Healthy class or Faulty class), One-class SVM for identifying an outlier (to omit this measurement from

S.V. Shrikhande (✉)
Software Development Section, Electronics and Instrumentation Group,
Bhabha Atomic Research Centre, Trombay, Mumbai, India
e-mail: svs@barc.gov.in

P.V. Varde
Research Reactor Services Division, Bhabha Atomic Research Centre,
Trombay, Mumbai, India
e-mail: varde@barc.gov.in

D. Datta
Radiological Physics and Advisory Division, Bhabha Atomic Research Centre,
Trombay, Mumbai, India
e-mail: ddatta@barc.gov.in

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_16

prognostic computation), Support Vector Regression, Fuzzy SVM (to give more weightage to recent data) and Kalman Filter for state prediction (for estimating remaining useful life) with uncertainty bounds.

Keywords SVM · SVR · FSVM · 1-Class SVM · LS-SVM · RUL

1 Introduction

Prognostics is anticipating impending faults and giving early warning of the failure before it happens. It is the process of predicting a product's Remaining Useful Life (RUL) by assessing the health degradation. Prognostics can be based on parameters correlated to degradation. By measuring such parameter(s) the remaining useful life can be predicted with uncertainty bounds. For CMOS ICs the quiescent power-supply current (I_{DDQ}) is the parameter which increases some orders of magnitude [2] and can be one of the factors used for prognosis. This has been stated in literature and also substantiated by carrying out Accelerated Life Testing (ALT) experiment [3]. These prognostics computations are done using soft computing techniques. This paper discusses soft computing techniques that can be used based on Support Vector Machines (SVM). This paper also discusses variants which are useful for different purposes viz. outlier identification, Support Vector Regression for forecasting and using recent data by employing Fuzzy Support Vector Machines and Least Squares SVM. This paper deals with their mathematical formulations and usage. For SVM variants, the optimization conditions are different and are covered in this paper. For estimating RUL, Support Vector Regression—a data driven technique and Kalman Filter—a model based technique for state prediction with uncertainty bounds is discussed.

The hybrid combining more than one technique—data-driven and model-based can be employed for prognostics.

2 Mathematical Methodologies for Prognostics

There are two principal methodologies to fault diagnosis and prognosis; one based on system identification and modelling and the other based purely on data-driven approaches using statistical computational intelligence techniques.

The system identification methodology will be tried to fit the model. Based on the Auto Correlation Factor (ACF) and Partial Auto Correlation Factor (PACF), the model will be identified. The main advantage of this approach is the ability to incorporate the physical understanding of the underlying process. On the other hand, data-driven approaches derive their information entirely from process data.

The drawback of such a method is its dependence on the quality of available process data.

Since degradation is a thermally governed process, after taking the natural logarithm the equation/model of the system becomes linear. This linearity is proved by the method of bi-coherency metric. Thus techniques of linear model becomes applicable.

Data driven Fault Detection techniques vary from simple threshold based fault detection to the sophisticated methods of Artificial Neural Networks (ANN), Fuzzy Logic approaches, Wavelet Analysis, Principal Component Analysis (PCA), Independent Component Analysis (ICA), etc.

As per literature survey, the ANN based approaches are used as fault classifiers both for binary fault classification as well as for multi-class fault classification. The major advantage of ANN is that theoretically they can approximate any continuous function without having any hypotheses of the underlying model. The disadvantage are that ANNs are like black-box where it is not possible to interpret the solution in traditional analytical way. Moreover the solution of ANN is not globally optimal and hence depends on the initial conditions of the network.

While ANNs are based on black-box technology, following a heuristic development with experimentation, the SVM are based on sound theory. SVMs are a learning tool based on statistical learning theory. SVMs have good generalization ability to unseen data. In the past few years SVM has shown excellent performance in may real world application including time series prediction. SVM are also suitable to data which is not regularly distributed or has unknown distribution. SVM involves finding solution to convex quadratic programming (QP) problem and gives unique global solution for positive definite kernel. This is advantageous as compared to ANN, which has multiple solutions associated with local minima and therefore not robust over different samples. By adaptively using the changes in the parameters, it is possible to prognosticate faults.

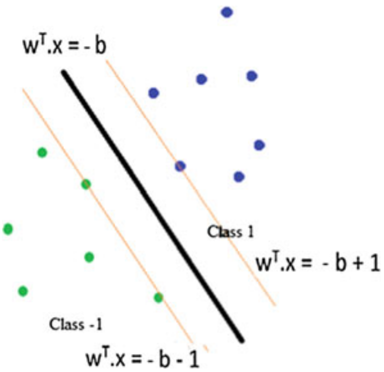
In model based approach like Kalman filter, the observation data is combined with pre-determined fault-growth model in order to update state predictions in an online manner.

3 Support Vector Machines

In machine learning, support vector machines are supervised learning algorithms that carry out pattern recognition from datasets i.e. it is a data driven technique. SVM are used for regression and classification.

SVMs have become one of the most popular approaches to learning from examples and have many potential applications in science and engineering. SVMs are relatively new computing methods. They are based on statistical learning theory, have high accuracy and show good generalization capability [5]. Also SVMs can handle data for any dimensionality.

Fig. 1 Hyperplanes of SVM



Using a labelled training dataset i.e., each data vector having its class label, an SVM training algorithm builds a model for hyper-plane. Though the simplest form of SVM is a linear binary classifier, it can also efficiently perform non-linear classification using kernel trick [4].

SVM solution is finding the hyperplane of (n-1) dimensionality for a general n-dimension problem; which is a line for 2-dimensional problem as shown in Fig. 1 and then draw two parallel hyperplanes to the hyperplane by pushing them as far apart as possible, until they hit data points. The classification plane with bounding planes furthest apart is the best one. Those points that touch the bounding plane, are called support vectors. The salient features are that all points in class 1 should be to the right of bounding plane 1 i.e.

$$\mathrm{w^T x_i} >= -\mathrm{b} + 1 \tag{1}$$

All points in class −1 should be to the left of bounding plane −1. i.e.

$$\mathrm{w^T x_i} <= -\mathrm{b} - 1 \tag{2}$$

Pick $\mathrm{y_i}$ to be +1 or −1 depending on the classification. Then the above two inequalities can be merged into one as given in Eq. (3).

$$y_i\left(w^T x_i + b\right) \geq 1 \tag{3}$$

The distance between bounding planes should be maximized. The distance between bounding planes is given by:

$$\frac{2}{\sqrt{w_1^2 + w_2^2 + \ldots + w_n^2}} = \frac{2}{\sqrt{w^T w}} \tag{4}$$

Thus the problem reduces to the convex optimization problem of

$$\min_{w,b} \frac{1}{2} w^T w$$

Such that

$$y_i \left(w^T x_i + b \right) \geq 1 \tag{5}$$

This SVM or hyperplane will be fitted based on the experimental data. SVM can accept as input, multiple features/condition indicators at the same time and produce as output the binary decision function.

Solving for finding the hyperplane is a mathematical optimization problem subject to constraints given in Eq. (5) [1]. More specifically, this is a quadratic solution problem and hence is a convex optimization problem.

3.1 Canonical SVM Problem

Since many solutions are possible by scaling w and b as stated above, we restrict our attention to a canonical solution (w, b) for which,

$$\min_i \frac{(w^T \phi(x_i) + b) y_i}{\|w\|} = 1 \tag{6}$$

So we get,

$$\begin{gathered} \max_w \frac{1}{\|w\|}, \text{ s.t.} \,\forall \text{i}, \\ \left(w^T \phi(x_i) + b \right) y_i \geq 1 \end{gathered} \tag{7}$$

Equivalent to the above equation is also the following equation

$$\begin{gathered} \min_w \|w\|^2, \text{ s.t.} \,\forall \text{i}, \\ \left(w^T \phi(x_i) + b \right) y_i \geq 1 \end{gathered} \tag{8}$$

The optimization problem of maximizing the margin can be brought down to the minimization problem of the term

$$\phi(w) = \frac{1}{2} ||w||^2 \tag{9}$$

The weight vector and bias vectors for the optimal separating hyperplane are found out as a quadratic optimization problem.

3.1.1 Non-separable Data

If a separating plane does not exist as shown in Fig. 1, then find the plane that maximizes the margin and minimizes the errors on the training points by taking original inequality with a slack variable to measure error.

For overlapping data the objective function is given below

$$\begin{gathered}\min_{w,w_0}\|w\|^2 + c\sum_i \xi_i \text{ s.t. } \forall \text{i}, \\ \left(w^T\phi(x_i) + b\right)y_i \geq 1 - \xi_i \\ \text{where } \forall \text{i } \xi_\text{i} > = 0\end{gathered} \tag{10}$$

In soft margin, we account for the errors. The above formulation is one of the many formulations of soft SVMs. In the above formulation, large value of C means overfitting.

Three types of points can be seen in Fig. 2. They are:

1. Correctly classified but $\xi_i > 0$ or violates margin
2. Correctly classified but $\xi_i = 0$ or on the margin
3. Incorrectly classified but $\xi_i > 1$

Where C is a positive number that is chosen to balance the two goals. To prevent from over-fitting with noisy data the slack variable ξ_i is introduced to allow some data points to lie within the margin, the constant C > 0 determines the trade-off between maximizing the margin and the number of training datapoints within that margin (and thus training errors). Those points that touch the bounding plane, or lie on the wrong side, are called support vectors.

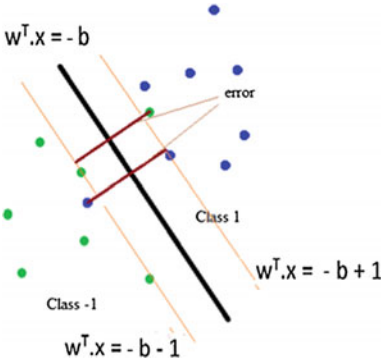


Fig. 2 SVM for non-separable data

3.1.2 Dual Formulation

Using the Lagrangian multipliers, the dual function is

$$\begin{aligned} & d^* = \max_{\lambda\varepsilon\Re}\ \min_{x\varepsilon D}\left(f(x) + \sum\nolimits_{i=1}^{m} \lambda_i g_i(x)\right) \\ & \mathrm{s.t.}\lambda_\mathrm{i}\ > \ = 0 \end{aligned} \tag{11}$$

The duality gap = 0 since the objective function as seen in Eq. (9) is a convex function. After solving the Kuhn Karush Tucker (KKT) conditions, we get **w** by the following equation.

$$w = \sum\nolimits_{i=1}^{m} \alpha_i^* y_i \phi^T(x_i) \tag{12}$$

To obtain w_o^* (or b), we can use the fact that, if $\alpha_\mathrm{i} \in (0, \mathrm{C})$, y_i (ϕ^T (x_j)w + w_0) = 1. Thus, for any point x_i such that, $\alpha_\mathrm{i} \in (0, \mathrm{C})$, that is, α_i is a point on the margin,

$$w_0 = \frac{1 - y_i\left(\phi^T(x_i)w\right)}{y_i} \tag{13}$$

$$\Rightarrow = y_i - \phi^T(x_i)w \tag{14}$$

The decision function,

$$f(x) = \phi^T(x_i)w^* + w_0^* \tag{15}$$

The classification function can then be written as:

$$\mathbf{f}(\mathbf{x}) = \mathbf{sgn}(\mathbf{f}(\mathbf{x}) \tag{16}$$

The bi-class classification of the online state between healthy state and faulty state will be using a bi-class SVM.

4 One-Class SVM

The primal one-class SVM problem for Novelty/anomaly detection is defined as

$$\begin{aligned} & \min_{w,\rho,e_i} \Im = \frac{1}{2} w^T w + \frac{1}{vn} \sum_{i=1}^{N} \xi_i - \rho \\ & \text{Subject to } w^T \phi(x_i) \geq \rho - \xi_i \\ & \xi_i \geq 0 \text{ for all i} = 1, \ldots, \mathrm{N} \end{aligned} \tag{17}$$

The parameter ν characterizes the solution:

It sets upper bound on the fraction of outliers (training examples regarded out-of-class).

It separates all the datapoints from origin and maximizes the distance from this hyperplane to the origin. Thus the function returns +1 in a small region and −1 elsewhere.

Like all other SVM formulations an equivalent dual problem is constructed and solved using Lagrange's multipliers as given in Eq. (18).

$$L(w, \rho, e_i, \alpha_i) = \Im - \sum_{i=1}^{N} \alpha_i \left(\left\langle w^T \phi(x_i) \right\rangle - \rho + \xi_i \right) \tag{18}$$

where αi are the Lagrangian multiplers, which can be +ve or –ve. Applying KKT conditions the conditions of optimality are-

$$\begin{aligned} \frac{dL}{dw} &=0 \Rightarrow w - \sum_{i=1}^{N} \alpha_i \phi(x_i) = 0 \\ \frac{dL}{d\rho} &=0 \Rightarrow \sum_{i=1}^{N} \alpha_i - 1 = 0 \\ \frac{dL}{de_i} &=0 \Rightarrow \alpha_i - \gamma e_i = 0 \\ \frac{dL}{d\alpha_i} &=0 \Rightarrow \sum_{i=1}^{N} w^T \phi(x_i) - \rho + e_i = 0 = 0 \end{aligned} \tag{19}$$

for k=1,…,N.

Putting these equations in matrix form-

$$\begin{bmatrix} I & 0 & 0 & -\phi \\ 0 & 0 & 0 & I \\ 0 & 0 & \gamma I & -I \\ \phi & -I & I & 0 \end{bmatrix} \begin{bmatrix} w \\ \rho \\ e \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \tag{20}$$

The matrix formalization can be written in a simplified form by eliminating w and e -

$$\begin{bmatrix} 0 & 1 \\ -1 & \Psi + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} \rho \\ \alpha \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{21}$$

where

$$\Psi_{kl} = \phi(x_k)^T \phi(x_l) \tag{22}$$

which is a kernel function.

This will be used for anomaly detection i.e. deciding whether the current input is an outlier. Outliers will not be considered for prognostic calculation and will be omitted.

The data obtained from experiment will be used as baseline data to construct a one-class SVM. This one-class SVM will be used with online data to distinguish between normal data which can be used for prognostics and abnormal/outlier data will be omitted from prognostics computation.

5 Fuzzy Support Vector Machine

Fuzzy Support Vector Machine (FSVM) is based on the idea that different input points can make different contributions to the learning of decision surface. Each input point is assigned a fuzzy membership, so that the different input points can make different contributions to the learning of decision surface. By setting different types of fuzzy membership, FSVMs can solve different kinds of problems. FSVMs can be used for Time series data, or for bi-class with different weightage or for reducing the effects of outliers. These are discussed below.

Time series data

For choosing the fuzzy membership function, choose the lower bound of fuzzy memberships ($\sigma > 0$). Time being the important factor, the fuzzy membership si can be made function of time ti

$$s_i = f(t_i) \tag{23}$$

where $t_1 \leq \ldots \leq t_l$ is the time the data arrived in the system.

We choose the latest data t_l as the most important and therefore

$$s_l = f(t_l) = 1 \tag{24}$$

Making the first point x_1 as the least important therefore

$$s_1 = f(t_1) = \sigma \tag{25}$$

If we fit a linear function of time for fuzzy membership then

$$s_i = f(t_i) = \alpha t_i + b \tag{26}$$

Applying boundary conditions, we get

$$s_i = f(t_i) = \frac{1-\sigma}{t_l - t_1} t_i + \frac{t_l\sigma - t_1}{t_l - t_1} \tag{27}$$

If we fit a quadratic function of time for fuzzy membership then

$$s_i = f(t_i) = \alpha(t_i - b)^2 + c \tag{28}$$

By applying boundary conditions, we get

$$s_i = f(t_i) = (1-\sigma)\left(\frac{t_i - t_1}{t_l - t_1}\right)^2 + \sigma \tag{29}$$

1-class with different weightage

Accuracy of one class is very high and of the other class is lower. For this the fuzzy membership is chosen as a function of the respective class. A sequence of training points are $(y_1, x_1, s_1), \ldots, (y_l, x_l, s_l)$

Fuzzy membership as function of class y_i is given by

$$\mathrm{s_i} = \mathrm{s}_+ \text{ if } \mathrm{y_i} = 1 \tag{30}$$

$$\mathrm{s_i} = \mathrm{s}_- \text{ if } \mathrm{y_i} = -1 \tag{31}$$

Typically $s_+ = 1$, $s_- = 0.1$

This fits the optimal hyperplane with errors appearing only in one class.

Reducing the effects of outliers by using class center

This can be done by setting the fuzzy membership as a function of the distance between the point and its class centre.

For the given sequence of training points, $(y_1, x_1), \ldots, (y_l, x_l)$.

Radius of class +1 is given by

$$r_+ = \max_{\{x_i : y=1\}} |\mu_+ - x_i| \tag{32}$$

where μ_+ is the mean of class +1. and Radius of class −1 is given by

$$r_- = \max_{\{x_i : y=-1\}} |\mu_- - x_i| \tag{33}$$

where μ_- is the mean of class −1.

Fuzzy membership s_i, a function of centroid and radius of class +1 and class −1 respectively are:

$$\begin{aligned} s_i &= \{1 - |x_+ - x_i|/(r_+ + \delta)\} \\ s_i &= \{1 - |x_- - x_i|/(r_- + \delta)\} \end{aligned} \tag{34}$$

where δ > 0 so that $\mathrm{s_i} \neq 0$. The distance of the two outliers to its corresponding mean is equal to the radii of the two classes. Due to the above fuzzy membership function, these two outliers are given least importance in FSVM training. This will cause a different hyperplane with reduced effect of outliers.

This will be used to reduce the intensive computations. The processing of large amount of historical data is drastically reduced using FSVM. The online time-series data which is acquired will be subjected to FSVM. The algorithm will be adaptive by giving higher weightage to recent data and progressively less weightage to older data. This technique will be used in alongwith SVR given below.

6 Least Squares SVM (LS-SVM)

This has low computational overhead. It does reduce accuracy but there is substantial gain in computation time and resources. In LS-SVM, the mathematical optimization equations are modified by adding a least squares term in the cost function. This eliminates solving QP problem and requires solution of a set of linear equations, thereby reducing the complexity of finding solution. In this technique the inequality constraints are changed to equality constraints. Implicitly the least square method is like regression.

Given training data $D = \left\{(x^1, y^1), \ldots.(x^N, y^N)\right\}$ with input data $x_i \in R^n$ and binary class labels $y \in \{-1, 1\}$, the LS-SVM is mathematically defined as the following optimization problem:

$$\min J_2(w, e) = \frac{\mu}{2} w^T w + \frac{\xi}{2} \sum_{i=1}^{N} e_i^2 \tag{35}$$

The above equation can be written as

$$\min J_2(w, e) = \frac{\mu}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^{N} e_i^2 \tag{36}$$

subject to

$$y_i(w\phi(x_i) + b)_i = 1 - e_i \quad \mathrm{i} = 1\ldots\mathrm{N}$$

where $\gamma = \frac{\xi}{\mu}$ is the tuning parameter (ratio of individual parameters μ and ξ for regularization and error resp.) Error variables allow some tolerance to misclassification.

Because of the equality constraint the Lagrangian dual formulation gives simplified linear programming solution given below.

$$\begin{bmatrix} 0 & -Y^T \\ Y & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{matrix} 0 \\ 1 \end{matrix} \tag{37}$$

where

$$\begin{aligned} &\Omega = Z^T Z = y_k y_l k(x_k, x_l), \\ &Z = \left[\phi(x_1)^T y_1; \ldots; \phi(x_N)^T y_N\right] \text{ and} \\ &\alpha = [\alpha_1; \ldots; \alpha_N] \end{aligned}$$

Thus the classifier is found by solving the linear set of above equations.

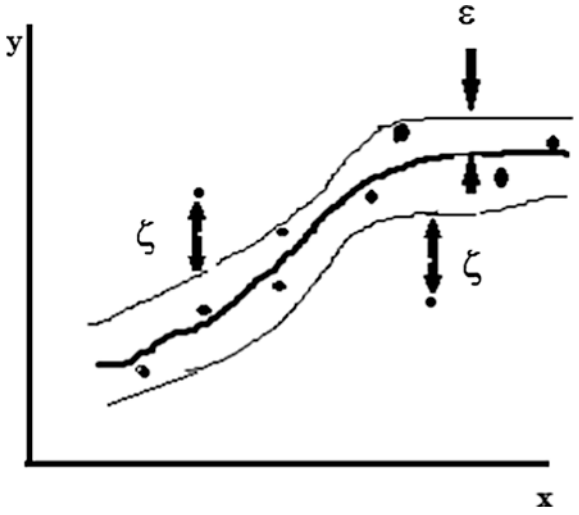
7 Support Vector Regression

Support Vector Regression (SVR) will be used for prediction of future data. The past time series data of a fixed duration will be used for finding the regression function.

The Support Vector method can also be applied to the case of regression (apart from classification problem), maintaining all the main features that characterise the maximal margin algorithm. This technique is useful more for regression problems, when sample data is sparse. A non-linear function is learned by a linear learning machine in a kernel-induced feature space while the capacity of the system is controlled by a parameter that does not depend on the dimensionality of the space. The Fig. 3 shows a situation for a non-linear regression function.

As long as points lie inside the ε margin, they do not contribute to the error. We can define the ε—insensitive loss function $L^{\varepsilon}(x, y, f)$ as given in Eqs. (38) and (39) and shown in Figs. 4 and 5 respectively.

Fig. 3 The insensitive band (slackness) for a non-linear regression function



For linear:

$$L^{\varepsilon}(x,y,f) = |y - f(x)|_{\varepsilon} = \max(0, |y - f(x)| - \varepsilon) \tag{38}$$

For Quadratic:

$$L_2^{\varepsilon}(x,y,f) = |y - f(x)|_{\varepsilon}^2 \tag{39}$$

In regression when slackness is introduced we have:-

$$\begin{aligned} &\min \frac{1}{2}\|w\|^2 + c\sum_i \xi_i^2 \\ &\text{s.t.}\ \forall \text{i}, \\ &\left(w^T\phi(x_i) + b\right) - y_i = \xi_i. \end{aligned} \tag{40}$$

The curve fitted by the regression technique of SVR is shown in the Fig. 6. The regression scheme associated with the latest data will be used for long-term predictions i.e. fault growth estimation with confidence bounds and remaining useful life (RUL) estimation after a fault is detected by subjecting to bi-class SVM.

Fig. 4 The linear ε—insensitive loss for zero and non-zero ε

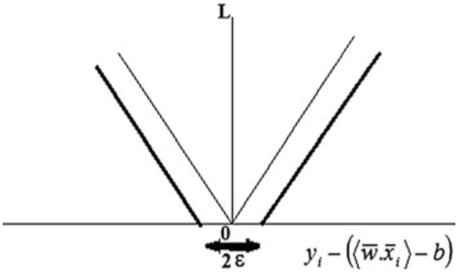
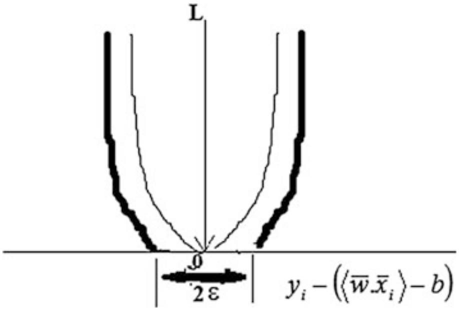


Fig. 5 The quadratic ε—insensitive loss for zero and non-zero ε



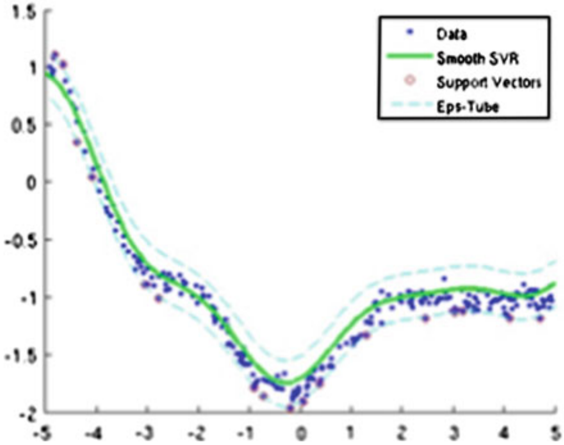


Fig. 6 Support vector regression [6]

8 Conclusion

SVMs are used as soft computing tools for various applications. They can be used effectively for prognostic computation. Its variants will be appropriately used like Bi-class SVM will be used to check whether the state is healthy or Failed. For the current input to check whether it is an outlier One-class SVM will be used. An outliers will not be considered for prognostic calculation and will be omitted.

Support Vector Regression (SVR) will be used for prediction of future data. The past time series data of a fixed duration will be used for finding the regression function. Fuzzy Support Vector Machine (FSVM) will be used alongwith SVR-it will be adaptive by giving higher weightage to recent data and progressively less weightage to older data. Least Squares SVM is a variant which has low computational overhead. Its effectiveness in terms of computational time reduction and the effect on reduction of accuracy will be tried out.

SVM alongwith its variants are powerful soft computing techniques for applications like prognosis.

References

1. Burges Christopher JC (2010) A tutorial on support vector machines for pattern recognition. Bell Laboratories, Lucent Technologies. Kluwer Academic Publishers, Boston (burges@lucent.com)
2. Roy K, Mukhopadhyay S, Meimand HM Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. In: Proceedings of the IEEE, vol 91, no 2, 20 Feb
3. Shrikhande SV, Varde PV, Datta D (2014) Identification and assessment of precursor in support of development of prognostics for CMOS ICs. In: National conference on reliability and safety engineering, pp 1–4. ISBN 978-8192112848

4. Shawe Taylor J, Cristianini N (2000) Support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge
5. Vapnik VN (1999) An overview of statistical learning theory. IEEE Trans Neural Networks 10:988–1000
6. Welling Max (2010) Support vector regression. www.ics.uci.edu/ ~ welling/teaching/KernelsICS273B/SVregression.pdf

Intelligent Real-Time Risk Analysis for Machines and Process Devices

Esko K. Juuso and Diego Galar

Abstract Automatic fault detection with condition and stress indices enables reliable condition monitoring to be combined with process control. Useful information on different faults can be obtained by selecting suitable features. Generalised norms can be defined by the order of derivation, the order of the moment and sample time. These norms have the same dimensions as the corresponding signals. The nonlinear scaling used in the linguistic equation approach extends the idea of dimensionless indices to nonlinear systems. The Wöhler curve is represented by a linguistic equation (LE) model. The contribution of the stress is calculated in each sample time, which is taken as a fraction of the cycle time. The cumulative sum of the contributions indicates the degrading of condition and the simulated sums can be used to predict failure time. To avoid high stress situations, the statistical process control (SPC) is extended to nonlinear and non-Gaussian data: the new generalised SPC is suitable for a large set of statistical distributions. It operates without interruptions in short run cases and adapts to the changing process requirements. The scaling functions are updated recursively, which is triggered by a fast increase of the deviation indices. The higher levels, which are rough estimates in the beginning, are gradually refined.

1 Introduction

Process control systems in industry include centralized or decentralized process controllers coupled with hosts, workstations and several process control and instrumentation devices, such as field devices. Applications are related to business

E.K. Juuso (✉)
Control Engineering, University of Oulu, P.O.Box 4300, 90014 Oulu, Finland
e-mail: esko.juuso@oulu.fi

D. Galar
Division of Operation and Maintenance Engineering, Luleå University of Technology,
Luleå, Sweden
e-mail: diego.galar@ltu.sel

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_17

functions in Enterprise resource planning (ERP) or maintenance functions in computerised maintenance management systems (CMMS). Smart field devices can include equipment monitoring applications which are used to help monitor and maintain the devices. A general architecture of IT systems related to operation and maintenance in process industry is shown in Fig. 1 [1].

The early detection of fluctuations in operating conditions and fault detection can be done with similar methods. Process and condition monitoring data is combined in detecting operating conditions (Fig. 2): normal process measurements are directly used in feature extraction, signal processing is needed for the condition monitoring data, and some infrequent measurements need to be interpolated [2]. Periodic condition monitoring measurements require interpolation to be used with other measurements in real-time systems.

Maintenance and operation performance are measured, monitored and analysed in many ways which provide information for the risk analysis [3]. Harmonised indicators can be used for monitoring maintenance actions on a management level, where the indicators are based on cost, time, man-hours, inventory value, work orders and cover of the criticality analysis, see [4]. Key performance indicators (KPIs), which focus on critical success factors and goals of the organization, differ depending on the organization [5]. Fully automated quantifiable KPIs would be

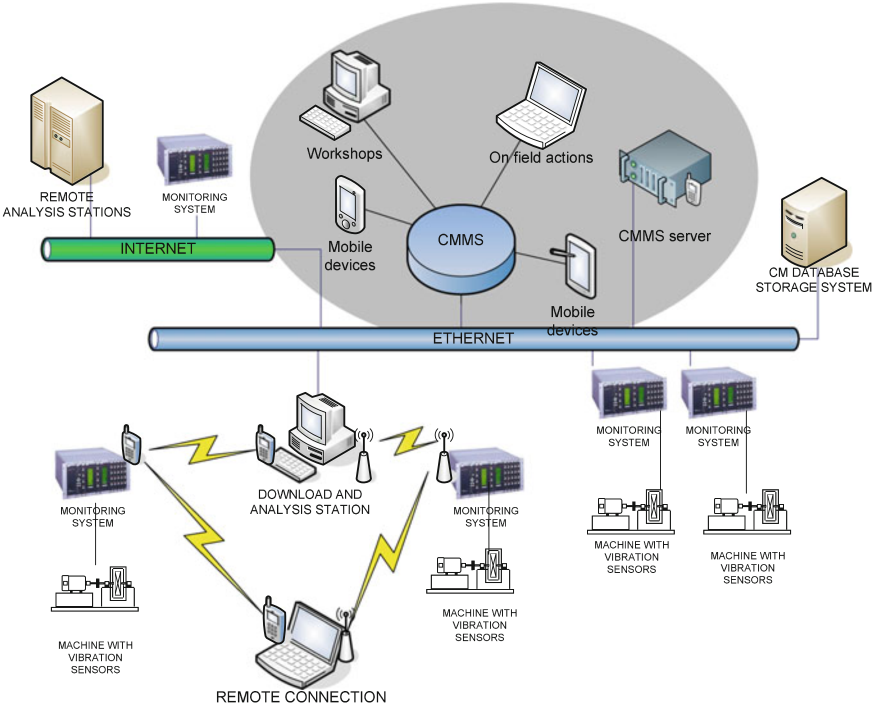


Fig. 1 Typical architecture of maintenance information system

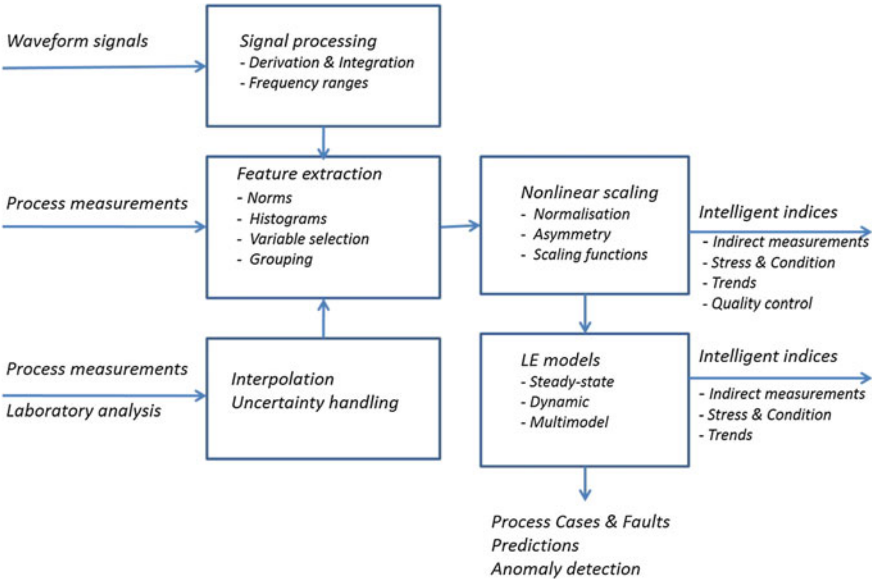


Fig. 2 Detecting operating conditions and faults

very useful. Overall equipment effectiveness (OEE) is a set of broadly accepted non-financial metrics which reflects manufacturing success [6]. Performance is evaluated by process capability indices (PCIs), which assume that process output is approximately normally distributed. Harmonised indicators, KPIs and PCIs can be handled as infrequent process measurements [3].

Real-time risk analysis requires more online measurements, where different wave form signals are important new sources of information. Vibration measurements provide a good basis for condition monitoring: elevated signal levels are detected in fault cases [7]. Mobile machines in underground mines introduce challenging environments for the measurements [8]. In rolling processes, torque is one of the most important measurements: the monitoring of rolling mill main drives requires torque to be measured directly at spindles or motor shafts since the main drives are highly dynamically loaded, which affect the product and the residual life time of drive components [9].

Efficient signal processing and feature extraction are essential in getting the waveform signals to real-time use: generalised norms [10] in wide variety of applications [11]. Linguistic equation (LE) models [12] based on nonlinear scaling [13] and data-driven tuning of the scaling functions [14] bring all the measurements to the same informative scale. Statistical process control (SPC), which was originally developed for quality control, is now widely used in fault detection and diagnostics [15]. Aggregation by features and indices is necessary for waveform signals. Nonlinear scaling is used in the same way for all the levels. Natural language can be used for all types of data [16].

This paper focuses on possibilities of the real-time risk analysis and feasible approaches to be used in the development. Signal analysis and feature extraction discussed in Sect. 2 are essential in intelligent analysers (Sect. 3) and in fatigue prediction (Sect. 4). Physical models are discussed in Sect. 5.

2 Signals and Features

Feature extraction uses derivation and statistical analysis. Real-time analysis can be based on the generalised norms [10], which has numerous applications [11]. Generalised spectral norms include the frequency domain in the time domain analysis [17].

2.1 Derivation

The calculation of the time domain signal $x^{(\alpha)}(t)$, which is based on a rigorous mathematic theory [18], is performed with three steps. The fast Fourier transform (FFT) is used for the displacement signal $x(t)$ to obtain the complex components $\{X_k\}, k = 0, 1, 2, \ldots, (N-1)$. The corresponding components of the derivative $x^{(\alpha)}(t)$ are calculated as follows:

$$X_{\alpha k} = (i\omega_k)^{\alpha} X_k, \tag{1}$$

where $\omega = 2\pi f$, $\alpha \in \Re$ is the order of derivation and f has integer values. Finally, the resulting sequence is transformed with the inverse Fourier transform FFT^{-1}, which produces the derivative signal. Since the vibration analysis is now based on the acceleration signals, the components of the derivative are obtained with an appropriate order of derivation $\alpha - 2$ [10].

2.2 Generalised Norms

The generalised norm defined by

$$\|{}^{\tau}M_{\alpha}^{p}\|_{p} = ({}^{\tau}M_{\alpha}^{p})^{1/p} = (\frac{1}{N}\sum_{i=1}^{N} |x_i^{(\alpha)}|^{p})^{1/p}, \tag{2}$$

where $\alpha \in \Re$ is the order of derivation, the order of the norm $p \in \Re$ is non-zero, τ is the sample time and N is the number of measurement values in the sample. The norm (2) includes the norms from the minimum to the maximum, which correspond the orders $p = -\infty$ and $p = \infty$, respectively. The norm values increase with increasing order. The computation of the norms can be divided into the computation of equal sized sub-blocks, i.e. the norm for several samples can be obtained as the norm for the norms of individual samples [10]. This which means that norms can be recursively updated.

2.3 Generalised Spectral Norms

Generalised spectral norms are calculated from the frequency spectrum by

$$\|X^{\alpha}\|_{p}= (\frac{1}{N}\sum_{k=1}^{N}\left|\{X_k\}^{(\alpha)}\right|^{p})^{1/p}, \tag{3}$$

where $\{ X_k\}$ is the sequence of complex numbers, representing different frequency components of the signal [18]. This kind of norm can be used, to provide for information about the change in signal in a certain frequency range or frequency ranges.

3 Intelligent Analysers

Intelligent condition and stress indices are calculated from these features by nonlinear scaling. The nonlinear scaling approach, which also uses the norms and moments, improves sensitivity to small fluctuations.

3.1 Nonlinear Scaling

The basic idea of the linguistic equation (LE) methodology is the nonlinear scaling developed to extract the meanings of variables from measurement signals. The scaling function scales the real values of variables to the range of [−2, +2] which combines normal operation [−1, +1] with the handling of warnings and alarms. The scaling function contains two monotonously increasing functions: one for the values between −2 and 0, and one for the values between 0 and 2 [12]. Constraints of the monotonous increase defined in [13] and the data-driven tuning of parameters introduced in [14] form the current design methodology. Knowledge-based information obtained from natural language is translated to the same value range [−2, 2] with the indices and indicators calculated from numerical values [16].

3.2 Stress Indices

Cavitation is detected well with the feature $\max(\left\|{}^{3}M_4^{2.75}\right\|)$, whose scaled value is the cavitation index $I_C^{(4)}$. The index levels shown in Table 1 are consistent with the vibration severity criteria defined in VDI 2056 [19]. Similar results obtained in a rolling mill are used for fatigue prediction, see Sect. 4.1 [20]. Effects of the stress are studied for a mobile machine in an underground mine [21].

3.3 Condition Indices

The norms $\max(\left\|{}^{15}M_4^{1}\right\|)$ and $\max(\left\|{}^{15}M_4^{4.25}\right\|)$ are highly sensitive to faulty situations in the supporting rolls of a lime kiln. The corresponding condition indices are consistent with the vibration severity criteria, which originate from VDI 2056 [19]. Research is continuing for a mobile machine in an underground mine, where condition indices are obtained repeatedly in similar steady operating conditions [21].

3.4 Generalised SPC

Statistical process control (SPC) is used in monitoring a process through the use of control charts [15]. SPC is extended to nonlinear with a large set of non-Gaussian statistical distributions. It operates without interruptions in short run cases and adapts to the changing process requirements. The approach has been tested in two application cases: a rolling mill and an underground load haul dump (LHD) machine [22].

Table 1 Cavitation index and vibration severity criteria [14]

Cavitation index	Cavitation level	Severity
$I_C^{(4)} < -1$	Cavitation-free	Good
$-1 \leq I_C^{(4)} < 0$	Short periods of weak cavitation	Usable
$0 \leq I_C^{(4)} < 1$	Short periods of cavitation	Still acceptable
$I_C^{(4)} \geq 1$	Cavitation	Not acceptable

3.5 LE Models

Linguistic equations (LE) models are linear equations

$$\sum_{j=1}^{m} A_{ij}X_j + B_i = 0, \tag{4}$$

where X_j is the linguistic level for the variable $j, j = 1\ldots m$. Each equation i has its own set of interaction coefficients $A_{ij}, j = 1\ldots m$. The bias term B_i was introduced for fault diagnosis systems. Various fuzzy models can be represented by means of LE models, and neural networks and evolutionary computing can be used in tuning. The methodology provides a flexible environment for fault diagnosis applications, software sensors, risk analysis and detection of sensor failures [2].

Nonlinear effects are handled with the scaling functions, i.e. intelligent indices can be used directly in the models to build more specific indices: two scaled features are combined in the lime kiln case [14]; cavitation indices were also earlier combinations of two scaled features [23].

4 Fatigue and Wear Prediction

Fatigue is caused by repeated loading and unloading. The mechanism proceeds through cracks formed when the load exceeded certain thresholds. Structures fracture suddenly when a crack reaches a critical size. Intelligent stress indices based on the nonlinear scaling provide good indicators of the severity of the load. The Wöhler curve is represented by a LE model,

$$I_S = \log_{10}(N_C), \tag{5}$$

where the stress index I_S can be a scaled value of stress or a scaled value of a generalised norm obtained from signals. The contribution of the stress is calculated in each sample time, which is taken as a fraction of the cycle time. The cumulative sum of the contributions indicates the deterioration of condition and the simulated sums can be used to predict the failure time [20].

4.1 Roller Mill

Torque measurements collected from a rolling mill have been used in the testing of the approach. The feature is a combination of two norms and the stress index is calculated from two scaled features obtained by using the nonlinear scaling

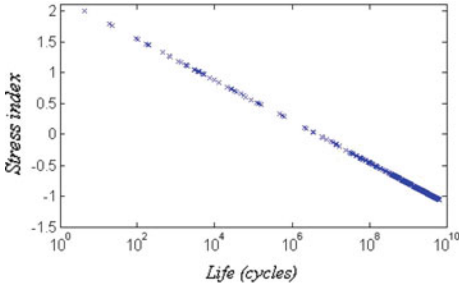


Fig. 3 Linguistic S-N curve presenting the analysed passes [20]

approach. The resulting linguistic S-N curve is linear (Fig. 3) and a normal S-N curve is formed from it by scaling to the feature values: a large number of passes have low stress indices. The high stress cases are seen as a very steep rise in the semilogarithmic curve [20].

4.2 Load Haul Dumper (LHD)

Cumulative stress analysis uses vibration measurements from the front axle of a load haul dumper (LHD). These machines operate in harsh conditions where failures may be difficult to repair. The machine is working in an underground mine. The cumulative stress increases fast during the high stress periods and increase is practically stopped when the stress is low since only stress indices are taken into account in the cumulative stress [21].

5 Maintenance

The described approach allows the prediction of system behaviour using either an analytical formulation of system processes (including degradation mechanisms) based on known principles or an empirically derived relationship. Many investigations into degradation mechanisms have been conducted, producing empirical damage models that are valid in a narrow range of conditions, such as wear, fatigue cracking and corrosion. Specific degradation mechanisms are generally studied and characterised under standard test conditions. Physics-based models are highly useful for describing the dynamics of time-varying systems, including different operating modes, transients, and variability in environmental stressors, but at the expense of the effort required to develop and validate the model [24].

The key challenge for a degradation model is to develop appropriate constitutive relationships for the condition decrease during degradation accumulation and to observe the complementary variables that characterise the relationship.

In the railway field, there are many physical models already validated that characterise the degradation of both track and rolling stock. For instance, the deterioration of track quality is often assumed to be proportional to the current quality. In this sense, a track in good condition deteriorates more slowly than a track in bad condition. This is usually modelled with the equation

$$Q(t) = Q_o * e^{bt}. \tag{6}$$

Here, Q_o is the track quality at time t = 0 and the parameter b is the deterioration rate characterising the behaviour in time. That is, the quality measure evolves according to an exponential model.

In Sweden, geometrical measurements are taken approximately every 1–2 months, excluding the winter. In Fig. 4, we fit the exponential model to the available data, ranging from April 2007 to September 2012 (21 geometrical measurements total). The exponential model, however, can only be fitted to deterioration branches with more than four measurements. However, the selected condition indicator is not enough for assets with repair and reliability restoration. In Fig. 4, the red lines represent maintenance actions on the track which eventually restore partially the reliability and certainly modify the condition indicator. These maintenance actions must be assumed as different since the maintenance performance always depends on many factors and the assumption of "as good as new" after intervention must be ignored.

As a summary, we find that this scenario of degradation combined with maintenance is a series of condition indicators modified by both mechanisms (degradation and restoration) and therefore the best tool for this modelling is time series [25].

Effects on railway data shown in Figs. 5 and 6 provide a more realistic description of the degradation in combination with maintenance, information that necessarily must be fused to get the holistic view of the asset management [26].

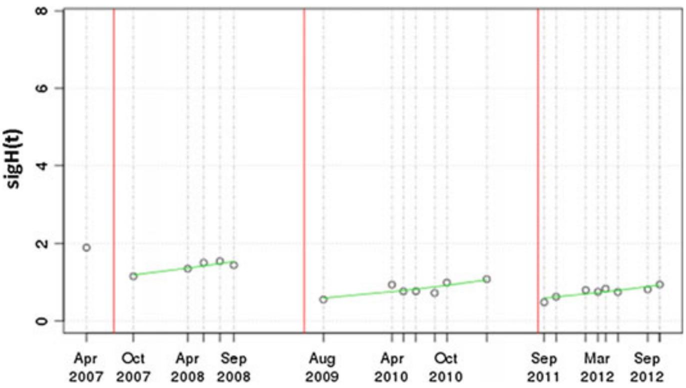


Fig. 4 Good fit of work orders and adjusted exponential model for sigH value

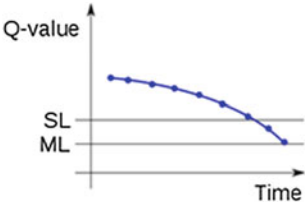


Fig. 5 Degradation of Q-value

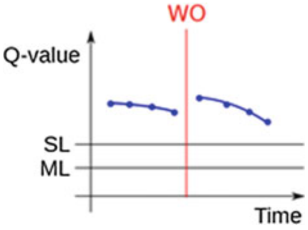


Fig. 6 Degradation followed by restoration of Q-value after a work order (WO)

The blue lines represent the exponential degradation models of the homogeneous sections, whereas the vertical (red) lines correspond to the work orders carried out in the homogeneous section. The thresholds SL and ML represent the Service Level and Maintenance Level, i.e. thresholds which require maintenance intervention if reached (proposed by the international standards) or levels of comfort worse than required in the SLA (Service level agreements).

The Q-value corresponds to the health index SOL, which can be calculated from the condition index by

$$SOL = 1 - \frac{2 - I_C^*}{4}(1 - \delta), \tag{7}$$

where δ is the value of SOL index when the condition index $I_C^* = -2$. In [27], SOL was calculated from the cavitation index, which is a stress index.

In summary, condition monitoring systems modelled as proposed in the previous chapter can be successfully combined with work order information from the CMMS by the means of using thresholds for features in time series data [28]. This assumption introduces in the model not only the degradation mechanisms but also the eventual restoration of the system.

6 Conclusions

Compact solutions have been developed for all the necessary steps of the real-time risk analysis: (1) features are specific for fault type and components, (2) intelligent analysers bring all the measurements and features to the same scale, (3) dynamic LE models based on intelligent indices provide predictions on fatigue and wear. All the methodologies are developed for calculations to be done in each sample time used in process control.

Acknowledgments The risk analysis approach has been developed within the research program “Measurement, Monitoring and Environmental Efficiency Assesment (MMEA)” on the basis of two projects “Production integrated condition-based maintenance model for mining industry (DEVICO)” and “Integrated condition-based control and maintenance (ICBCOM)”. The funding of the TEKES (the Finnish Funding Agency for Technology and Innovation) and several companies is acknowledged.

References

1. Galar D, Kumar U, Juuso E, Lahdelma S (2012) Fusion of maintenance and control data: a need for the process. In: Proceedings of WCNDT, 16 pp
2. Juuso E, Leiviskä K (2010) Combining process and condition monitoring data. In: Maintenance, condition monitoring and diagnostics, proceedings of the 3rd international seminar, POHTO, Oulu, pp 120–135
3. Juuso EK, Lahdelma S (2013) Intelligent performance measures for condition-based maintenance. J Qual Maint Eng 19(3):278–294. doi:10.1108/JQME-05-2013-0026
4. Olsson C, Svantesson T (2009) Harmonised maintenance and reliability indicators—compare apples to apples. Maintworld 1(1):9–11
5. Al-Shammasi NA, Al-Shakhoyry SS (2010) Improving maintenance performance in Saudi Aramco. Maintworld 2(2):6–9
6. SCEMM (1998) SCEMM keep it running—industrial asset management. Painoyhtymä, Loviisa
7. Lahdelma S, Juuso E (2007) Advanced signal processing and fault diagnosis in condition monitoring. Insight 49(12):719–725
8. Laukka A, Saari J, Ruuska J, Juuso E, Lahdelma S (in press). Condition based monitoring for underground mobile machines. Int J Ind Syst Eng Spec Issue Maint Perform Meas Manag
9. Mackel J, Fieweger M (2010) Condition monitoring in steel industry. In: Maintenance, condition monitoring and diagnostics, proceedings of the 3rd international seminar, POHTO, Oulu, pp 26–47
10. Lahdelma S, Juuso E (2011) Signal processing and feature extraction in vibration analysis, part I: methodology. Int J Cond Monit 1(2):46–53. doi:10.1784/204764211798303805
11. Lahdelma S, Juuso E (2011) Signal processing and feature extraction in vibration analysis, part II: applications. Int J Cond Monit 1(2):54–66. doi:10.1784/204764211798303814
12. Juuso EK (2004) Integration of intelligent systems in development of smart adaptive systems. Int J Approx Reason 35(3):307–337. doi:10.1016/j.ijar.2003.08.008
13. Juuso EK (2009) Tuning of large-scale linguistic equation (LE) models with genetic algorithms. Adaptive and natural computing algorithms, Revised selected papers—ICANNGA 2009, Kuopio, Finland ICANNGA 2009. Lecture notes in computer science (LNCS) 5495. Springer, Heidelberg, pp 161–170. doi: 10.1007/978-3-642-04921-7_17

14. Juuso E, Lahdelma S (2010) Intelligent scaling of features in fault diagnosis. In: Proceedings of CM 2010/MFPT 2010, BINDT, vol 2, pp 1358–1372. www.scopus.com
15. Oakland JS (2008) Statistical process control, 6th edn. Butterworth-Heinemann, Oxford
16. Juuso EK (2012) Integration of knowledge-based information in intelligent condition monitoring. In: Proceedings of CM 2012/MFPT 2012, vol 1, pp 217–228. www.scopus.com
17. Karioja K, Juuso E (2015) Generalised spectral norms—a new method for condition monitoring. In: Proceedings of CM 2015/MFPT 2015, pp 636–643
18. Samko SG, Kilbas AA, Marichev OI (1993) Fractional integrals and derivatives. Theory and applications. Gordon and Breach, Amsterdam, 976 pp
19. VDI 2056 Beurteilungsmaβstäbe für mechanische Schwingungen von Maschinen. VDI-Richtlinien, Oktober 1964
20. Juuso E, Ruusunen M (2013) Fatigue prediction with intelligent stress indices based on torque measurements in a rolling mill. In: Proceedings of CM 2013 and MFPT 2013, vol 1, pp 460–471. www.scopus.com
21. Juuso EK (2014) Intelligent indices for online monitoring of stress and condition. In Proceedings of CM 2014/MFPT 2014, vol 1, pp 637–648. www.scopus.com
22. Juuso EK (2015) Generalised statistical process control (GSPC) in stress monitoring. In Proceedings of IFAC MMM 2015, p 6
23. Juuso E, Lahdelma S (2006) Intelligent cavitation indicator for Kaplan water turbines. In Proceedings of COMADEM 2006, pp 849–858
24. Galar D, Kumar U, Villarejo R, Johansson CA (2013) Hybrid prognosis for railway health assessment an information fusion approach for PHM deployment. In: 2013 Prognostic and system health management PHM 2013, Milan, 8–11 Sept 2013
25. Pajares RG, Benítez JM, Palmero GS (2008) Feature selection for time series forecasting: a case study. In: Eighth international conference on hybrid intelligent systems HIS'08, IEEE, pp 555–560)
26. Galar D, Gustafson A, Tormos B, Berges L (2012) Maintenance decision making based on different types of data fusion. Eksploatacja i Niezawodnosc Maint Reliab 14(2):135–144
27. Juuso E, Lahdelma S (2010) Intelligent condition indices in fault diagnosis. In: Proceedings of CM 2008, BINDT, pp 698–708
28. Galar D, Thaduri A, Catelani M, Ciani L (2015) Context awareness for maintenance decision making: a diagnosis and prognosis approach. Measurement 67:137–150

Part III
Maintenance Management

Malfunction in Railway System and Its Effect on Arrival Delay

Sida Jiang and Christer Persson

Abstract In this research project, we have established a set of advanced statistics models that quantify the cause-effect relation between infrastructure failures and train delay. The major model we employed in this project is called the "Wiener process model", and we are the first researching team to implement the Wiener-process model into traffic analysis area with large scale network data in Swedish railway. The data we based our research on includes a 1) train movement record database—TFÖR 2) infrastructure error reporting system—0FELIA 3) railway facility database—BIS. For TFÖR alone, there is a 27-million data record over 5 different rail classes (from rail class 1, major railway around big city areas to rail class 5 least loaded rail) and 3 different passenger train types (x2000, regional train and commuter train). By merging the database listed above, a specified wiener process model has been estimated for the primary delay caused by system errors and the secondary delay by interaction of trains. The model also quantifies the effects of characteristics of railway system over different rail classes and operation manners. In addition, the Wiener process model also enable further research to derive the fundamental relation between capacity, speed and density (inverse function of time gap) in railway context.

Keywords Cause-effect relation · Wiener process · System errors · Arrival delay and Meso-level simulation

S. Jiang (✉)
WSP Analysis & Strategy, Arenavägen 7, Stockholm-Globen, Sweden
e-mail: sida.jiang@wspgroup.se

C. Persson
KTH, Teknikringen 72, Stockholm-Globen, Sweden
e-mail: christer.persson@abe.kth.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_18

1 Introduction

The railway system is a complex system and errors that occur in the system can cause delay for the passengers travelling by train. However knowledge of the relationship between railway errors and travel delay, in the current situation is relatively deficient. The aim of the project is to put forward the knowledge by estimating the relationship between infrastructure-related errors that occur in the system and resulting delay in railway.

The alternative solution available today is basically micro-level simulation, e.g. by means of RailSys [1], to simulate how the system failure affects the delays in train traffic. In the current situation, this is an option only for certain individual sections of the railway system due to the regard of data complexity and the long simulation time. Besides requiring detailed features of the studied railway network, self-defined delay distribution needs also to be constructed in Railsys simulation [2]. Wiener process, on the other hand, is a statistical model to quantify cause-effect relationship between errors and delays that can be applied to general analytical scenarios, such as varied time table frequency, real-time headway and track type. The arrival delay distribution derived from the proposed Wiener process can also lay out a basis for delay distribution input to Railsys and other micro-simulation tools in railway research.

In this study, rather than a general model that covers the whole railway system, the individual trains are described individually in the model, but far less detailed than in standard simulation model. The level of details to predict the effects of individual trains will be less than in the simulation model, but the opportunities to analyze the majority or large parts of the system increases. The chosen model estimates the travel time probability distribution for individual trains and route segments. The transport model at this context is called the meso-level simulation, while usual practice uses the RailSys system at the micro level.

The model used is a so-called stochastic diffusion. It is a flexible model that contains great customization options which can be applied in the sector. In this project, the main purpose was to find out in what extent has break down of the railway system affected arrival delay, and how to capture primary and secondary delay in the Wiener process model, without regard to details of the whole railway system.

2 Definition and Denotation

2.1 *The Wiener Process*

The Wiener Process (W) is frequently referred as Brownian motion. One way to modify the Wiener Process in the context of train movement is to introduce a systematic drift (μ) in which the train moves towards the destination while the infrastructure failures yield the diffusion (diffusion coefficient σ) against the direction of the drift. If the train starts at station say $W(0) = d$ with distance d to the next station and it takes time T—often termed as first passage time or hitting time, thus $T = inf_{t \geqq 0}\{t{:}W(t) = 0\}$. The distribution of T is an inverse Gaussian distribution and the process $X(t) = d - \mu t + \sigma\, W(t)$ is called a Wiener process [3]. The Wiener process has been frequently applied in pure & applied mathematics, physics, economics and quantitative finance (in particular the Black-Scholes option pricing model). The application of the results of "hitting time" to multiple channel queues in heavy traffic is firstly discussed by Lglehart and Whitt [4].

2.2 Arrival Delay

In stochastic process theory, the Wiener Process is featured with its stationary and independent increments which give great convenience in practicing frailty theory in railway context; nonetheless the train movement is decomposed into a set of independent movements over railway links between two consecutive stations/stops. In order to avoid the violation of the independency of the decomposing train movement over different links, we thereby introduce ***arrival delay*** as shown in following formula [5] (Fig. 1):

$$Arrival\ delay = arrival\ time - planned\ arrival\ time - departure\ delay \qquad (2.1)$$

So that the defined arrival delay is independent of accumulated delay from previous links, in the same manner that the probable arrival delays of links afterwards is also independent of the studied link. The study focuses on the passenger traffic and the station pairs that have the correct record of arrival time.

2.3 Primary and Secondary Delay

In railway operation, train delay is usually categorized into primary and secondary delay due to its causes. Primary delay is majorly caused by malfunctions in

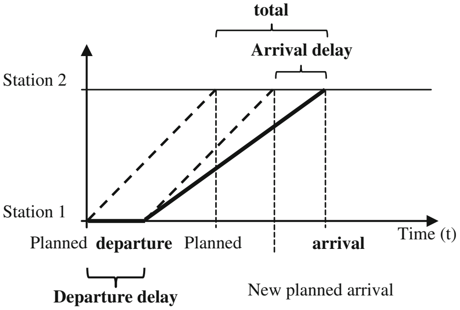


Fig. 1 Diagram of relation between arrival delay, arrival time, planned arrival time and departure delay

infrastructure (signal, track, communication etc.) or train. The secondary delay is known as "knock-on" delay in the sense that delays may be propagated to other trains due to interaction between trains [6]. In order to improve the robustness of the real-time railway operations, buffer time is supplemented in the time table to absorb the delay caused by both primary and secondary delay. The optimization of time table with regards to time supplement has frequently discussed in railway simulation and practice [7], yet the system malfunction and interaction between consecutive trains is more stochastic than the planned time table, therefore in this study we have also investigate the real-time time gap between consecutive trains instead of supplement time in time table.

All the errors of the Swedish railway system is registered in a database called 0FELIA. To derive the cause-effect relation between railway system and train arrival delay, we have firstly matched 0FELIA with a train movement database called TFÖR using the shortest path method. Hence, each row of the combined dataset is an individual train movement with system error(s) that occurred at the same link and can probably affect the train. In addition, the closest train movement ahead of the studied train movement (upon the same link) is also attached in the same row of combined dataset. To model how the interaction between consecutive trains yields secondary delay, we also computed the headway or time gap between closest train pairs. Our combined dataset thereby integrates train movement information with affecting system errors and the headway to the train ahead.

3 Data and Method

3.1 Overview of the Model

Each and every individual train movement between two train stations/stops can be specified in the following Wiener process model [3]:

$$X(t) = d - \mu t + \sigma W(t) \tag{3.1}$$

where

t is time;

$X(t)$ is the position of the train from the starting station along the link measured in kilometer. Eg. $X(0) = d$

$W(t)$ is a standard Wiener process, Var$W(t) = t$

d is the length of the link (kilometer), $d > 0$

μ is the average speed that train moves over the link

σ is the variation of the location of train (Var$X(t) = \sigma^2 t$)

With *d* and *μ > 0* yields a negative sign for *μ* in (3.1) such that the train started from $X = d$ to the station along the link at the position $X = 0$.

The arrival time T_a in the terminology of wiener process is called "first hitting time at 0", and can be formed as follows:

$$T_a = \inf\{t : X(t) = 0\} \tag{3.2}$$

The arrival time T_a follows the inversed Gaussian distribution and can be estimated through maximum likelihood. Two important features can be derived from the Wiener process model for average travel time E(T_a) and the variation of travel time Var(T_a) over the studied link:

$$E(T_a = \frac{d}{\mu}) \tag{3.3}$$

$$Var(T_a) = \frac{d\sigma^2}{\mu^3} = E(T_a)\frac{\sigma^2}{\mu^2} \tag{3.4}$$

The travel time and its variation are thus easily estimated through the average speed and variation of train location from the standard Wiener process model. Important characteristics of the railway system have been modelled as the variant of the average speed μ and location variation σ. Assuming Z is the independent variable matrix including majorly:

- Rail class
- Train type
- System errors
- Time gap

Then

$$\mu = Z\beta \tag{3.5}$$

$$\sigma = \exp\{Z\delta\} \tag{3.6}$$

Exponential transformation is used to assure the positive sign of σ. β represent the coefficients for independent variables and is estimated through maximum-likelihood, through coefficients we can further calculate e.g. the effects of changing corresponding railway characteristics upon the average speed and arrival delay.

3.2 Model Specification

Specification of average speed and variation component is formulated with a series of variants that represent the characteristics of the corresponding railway system:

$$\begin{aligned}\mu = {} & \beta_1 + \beta_2 \mathit{Rail\ Class}2 + \beta_3 \mathit{Rail\ Class}3 \\ & + \beta_4 \mathit{RailClass}4 + \beta_5 \mathit{RailClass}5 + \beta_6 \mathit{RailClassNA} + \beta_7 x2000 + \beta_8 \mathit{CommuterTrain} \\ & + \frac{\beta_9}{\mathit{timegap} + 1} + \beta_{10} \mathit{TotalErrors} + \beta_{11} \mathit{TotalErrors} \cdot \mathit{RailClass}4 \\ & + \beta_{12} \mathit{TotalErrors} \cdot \mathit{RailClassNA} + \frac{\beta_{13} \mathit{RailClas}2}{\mathit{timegap} + 1} + \frac{\beta_{14} \mathit{RailClas}4}{\mathit{timegap} + 1}\end{aligned} \tag{3.7}$$

Constant β_1 is user specified and in the formula above it represents the reference rail class and train type, that is rail class 1 and regional passenger train. Rail class NA is one small part of data that has missing values for rail class information. Total errors are the system errors registered in 0FELIA that occurred yet not fixed at the arrival station when the train starts. Alternatively, system malfunction can be categorized into errors of different types: rail, signal, communication, electricity etc., but we only use the number of errors since certain error category may not satisfy the required minimum size due to rare occurrence.

The proposed hypothesis for the effect of the time gap upon the secondary delay is up to certain limit, the marginal effect of one unit time-gap increment is diminished over its magnitude. Therefore, the inverse function is introduced to describe its non-linear effect. The non-linear transformation has also been tested significantly to improve the goodness of fit for the model.

The specification of variation of location σ basically follows the same design as the average speed; nonetheless we applied an exponential function to assure the positive sign of σ:

$$\begin{aligned}\sigma = \exp\{ & \beta_{15} + \beta_{16} \mathit{RailClass}2 + \beta_{17} \mathit{RailClass}3 \\ & + \beta_{18} \mathit{RailClass}4 + \beta_{19} \mathit{RailClass}5 + \beta_{20} \mathit{RailClassNA} + \beta_{21} x2000 \\ & + \beta_{22} \mathit{CommuterTrain} + \frac{\beta_{23}}{\mathit{timegap} + 1} + \beta_{24} \mathit{TotalErrors} \\ & + \beta_{25} \mathit{TotalErrors} \cdot \mathit{RailClass}2 + \beta_{26} \mathit{TotalErrors} \cdot \mathit{RailClass}3 \\ & + \beta_{27} \mathit{TotalErrors} \cdot \mathit{RailClass}4 + \beta_{28} \mathit{TotalErrors} \cdot \mathit{RailClass}5 \\ & + \beta_{29} \mathit{TotalErrors} \cdot \mathit{RailClassNA} + \frac{\beta_{30} \mathit{RailClas}2}{\mathit{timegap} + 1} + \frac{\beta_{31} \mathit{RailClas}4}{\mathit{timegap} + 1}\}\end{aligned} \tag{3.8}$$

More combined total errors with rail class in the formulation of location variation are proved significantly.

3.3 Data and Sampling

There are 27.6 million train movements for 2009 in the TFÖR database, of which contains 18.2 million records for passenger traffic. Each train movement record is a combination of train number, train type, rail class (bantyp in Swedish) and travel time related. The large data size motivates a statistical sampling through TFÖR then combined with 0FELIA and time gap to corresponding closest train.

In the following tables, two ways to look at the "representativeness" of passenger train movement over Sweden in 2009 from TFÖR database are presented (Tables 1 and 2):

The sample containing 90 000 passenger train movement is selected in the project to represent the identical structure of different rail classes in Sweden, 2009, while with 100 unique train numbers for each train type, the weights of different train type is computed for further modal adjustment and the interpretation of results. After data processing we have obtained 35 596 observations as our input data.

4 Results and Summary

4.1 Estimation Results

The log-likelihood of the model is −77 023.76 with total 35 596 observations. By average, x2000 on rail class 2 has the highest speed while rail class 5 or commuter train in rail class 1 has significantly lower speed in the Swedish railway network. For primary delay: total errors across all rail classes have negative effects upon average speed, yet for rail class NA even more negative than any other rail class which may be quite meaningful to identify its rail class in further studies. For

Table 1 Division of passenger train movement over different train types in TFÖR, 2009

Train type	Share (%)
X2000	12
Commuter train	42
Other trains	46

Table 2 Division of passenger train movement over different rail classes in TFÖR, 2009

Rail class	Share (%)
1	37.7
2	34.9
3	21.4
4	5.7
5	0.4
In total	100.0

Note Rail class 1—metropolitan areas e.g. Stockholm to 5—the least traffic loaded rail class

Table 3 Estimation results for passenger train sample

Coeff.	Parameter	Value	P-value
Average speed μ			
β1	Intercept	1.73	<0.0001
β2	RailClass2	0.32	<0.0001
β3	RailClass3	0.04	<0.0001
β4	RailClass4	0.02	0.6068
β5	RailClass5	−0.61	<0.0001
β6	RailClassNA	0.38	<0.0001
β7	x2000	0.46	<0.0001
β8	Commuter Train	−0.09	<0.0001
β9	1/(timegap + 1)	−0.68	<0.0001
β10	Total errors	−0.16	<0.0001
β11	RailClass4*total errors	0.10	0.0266
β12	RailClassNA*Total errors	−0.02	<0.0001
β13	RailClass2* (1/(time gap+1))	0.38	<0.0001
β14	RailClass4* (1/(time gap+1))	−0.37	0.0007
Variation of location σ			
β15	Intercept	0.18	<0.0001
β16	RailClass2	0.24	<0.0001
β17	RailClass3	−0.07	<0.0001
β18	RailClass4	0.59	<0.0001
β19	RailClass5	−0.16	0.6886
β20	RailClassNA	0.13	<0.0001
β21	x2000	0.11	<0.0001
β22	Commuter train	−0.19	<0.0001
β23	1/(timegap+1)	0.33	<0.0001
β24	Total errors	−0.06	<0.0001
β25	RailClass2* total errors	0.11	<0.0001
β26	RailClass3* total errors	0.13	<0.0001
β27	RailClass4* total errors	−0.22	0.0004
β28	RailClass5* total errors	1.05	0.0321
β29	RailClassNA*total errors	−0.06	<0.0001
β30	RailClass2*(1/(time gap+1))	−0.83	<0.0001
β31	RailClass4*(1/(time gap+1))	−2.56	<0.0001

secondary delay: significant effects of time gap has been identified for rail class 2 & 4 with opposite signs compared with rail class 1, but in general the time gap has, *at different extent*, a positive effect to increase the mean speed. An extra time is often added in the design of the time table to diminish the potential secondary delay, more reliability can be gained through a time gap plug-in in rail class 4 since it is usually single track and less maintained. Certainly, the fundamental relation between A) traffic flow (indicated by rail class) & mean speed B) traffic flow & density (can be formed by time gap) is also non-linear in railway context. It needs to

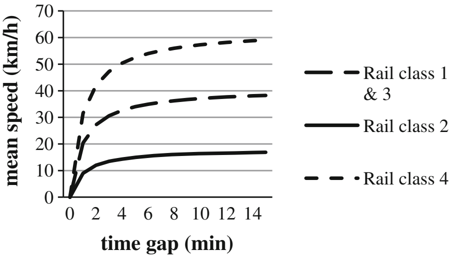


Fig. 2 The effects of time gap upon mean speed over different rail classes

be explored further through different modelling approaches, here is just a priori variant (Rail Class X* (1/(time gap+1))) to test the significance of a combined variable as well as the cause-effect relation between time gap and secondary delay. The coefficient estimation can be seen in Table 3:

In Fig. 2, a comparative result has been analyzed and illustrated where we can find that the time gap has in general an "approximately monotone" positive effects upon corresponding mean speed, yet all the effects is converged to certain speed level for different rail classes. Furthermore, the diminishing marginal curve means that it is more effective to plug in short unit of time in heavily loaded traffic situation than otherwise. Rail class 1 & 3 been affected most since rail class 1 has quite intensive train operations, also with a mixture of different train type at large extent; rail class 3 is almost all single track and to great extent subjected to the minimum time gap from both directions.

The parameter σ describes the variation of train location; a higher value yields higher variation of travel time. The effect upon variation is an exponential function of estimated coefficients. We have noticed that the sign for time gap inversion in rail class 1 is positive which means with time gap increasing the variation will diminishes in the amount of $\exp(0.33) \approx 1.39$, in a way to reduce the unreliability of travel time. Again the combination of rail class and time gap needs to be further adjusted so that it will be comparable with fundamental relation in roadway.

4.2 Summary

The research has specified the Wiener process model and implemented it into a large scale simulation in the Swedish railway system. The merging of train movement database TFÖR and system error registry 0FELIA together with data processing for important variable such as time gap, has enabled the railway administrator to investigate at the meso-level both primary and secondary delay with respect to a series of important characteristics over railway system. To enrich the understanding of the fundamental relation between capacity, speed and density (inverse function of time gap) in railway context, future research needs to firstly build upon more updated and detailed database e.g. Lupp such that track information for train movement and travel time in seconds can be employed. Furthermore, the application of the research results is not limited to time table design or railway operation to minimize travel delay; we also foresee strong motivation to calculate the elasticity as a basis for socio-economic effects of different maintenance strategies.

Acknowledgments We would like to thank our sponsors of ICRESH-ARMS 2015 for their intellectual and financial support. Special thanks to Joel Åkesson at Trafikverket, Håkan Berell at WSP, Pia Sundbergh at Trafa and Dr. Alireza Ahmadi at LTU to contribute their instructive insights and generous help.

References

1. Bendfeldt J-P, Mohr U, Müller L (2000) RailSys, a system to plan future railway needs. International conference on computers in railway, Bologne
2. Sipilä H (2012) Simulation of rail traffic—Applications with timetable construction and delay modelling. Licentiate thesis, KTH, Sweden
3. Aalen O, Borgan Ø, Gjessing HK (2008) Survival and event history analysis. Springer, New York
4. Iglehart DL, Whitt W (1970) Multiple channel queues in heavy traffic. I Adv Appl Probab 2 (1):150–177
5. Jiang S, Persson C, Sundbergh P (2012) Fel i järnvägssystemet och dess effekter på förseningar, working report, WSP
6. D'Ariano A, Pacciareli D, Pranzo M (2007) A branch and bound algorithm for scheduling trains in a railway network. Eur J Oper Res 183(2007):643–657
7. Kroon L, Maróti G, Helmrich MR et al (2008) Stochastic improvement of cyclic railway timetables. Trans Res Part B Methodol 42(6):553–570

On-Condition Parts Versus Life Limited Parts: A Trade off in Aircraft Engines

Veronica Fornlöf, Diego Galar, Anna Syberfeldt and Torgny Almgren

Abstract Maintaining an aircraft engine is both complex and time consuming since an aircraft is an advanced system with high demands on safety and reliability. Each maintenance occasion must be as effective as possible and the maintenance need to be executed without performing excessive maintenance. The aim of this paper is to describe the essence of aircraft engine maintenance and to point out the potential for improvement within the maintenance planning by improving the remaining life predictions of the On-Condition parts, i.e. parts that are not given a fixed life limit.

Keywords Aircraft engine maintenance · Remaining useful life · Reliability · On-Condition parts

1 Introduction

Aircraft engines are one of the most critical parts of an aircraft and are therefore where most of the maintenance efforts are allocated.

Efficient maintenance of an aircraft focus on how to ensure the realization of the inherent safety and reliability levels of the aircraft, and also to restore safety and

V. Fornlöf (✉)
University of Skövde, SE- 461 81, Trollhättan, Sweden
e-mail: veronica.fornlof@his.se

D. Galar
Luleå University of Technology, SE-971 87, Luleå, Sweden
e-mail: diego.galar@ltu.se

A. Syberfeldt
University of Skövde, Box 408 SE-541 28, Skövde, Sweden
e-mail: anna.syberfeldt@his.se

T. Almgren
GKN Aerospace Engine Systems, SE- 461 81, Trollhättan, Sweden
e-mail: torgny.almgren@gknaerospace.com

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_19

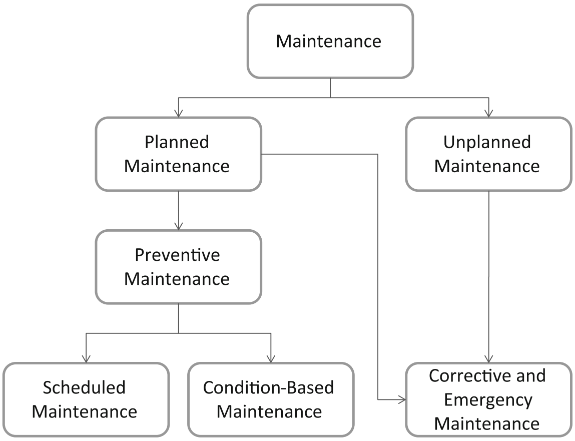
reliability to their inherent levels when deterioration has occurred [1]. Aircraft maintenance does also occupy a key position in airline operation because maintenance is essential to the safety of the passengers and the reliability of airline schedules [2]. An unexpected failure that could lead to an aircraft crash must be avoided by all available means. Maintenance, and to perform correct maintenance, is therefore a prerequisite for a successful aviation industry.

Maintenance is the combination of all technical and associated administrative actions intended to retain an item in, or restore it to, a state in which it can perform its required function. The goal is to prevent fatal damage for machine, human or environment and to prevent unexpected machine failure by using condition based maintenance planning to increase safety of production and quality control. Figure 1 below shows a breakdown of different maintenance strategies.

Basically there are three different maintenance strategies [3]:

- *Run-to-break* is the most simple maintenance strategy that is often used for systems that are cheap and where damage does not cause other failures. The machine or system is used until it breaks. It is commonly used for consumer products.
- *Preventive Maintenance* is the most common maintenance method for industrial machines and systems. With this strategy maintenance is performed in fixed intervals. The intervals are often chosen so that only 1–2 % of the machine will have a failure in that time.
- *Condition-Based Maintenance* is also called predictive maintenance. Maintenance is dynamical planned based on machine or system condition. Condition-Based Maintenance does have advantages compared to the other two strategies, since modern measurements and signal-processing methods are used to accurately diagnose item/equipment during operation. It though requires a reliable condition monitoring method. One area within this part of maintenance is condition monitoring which aims to continuously observe wear-related variables throughout a system's lifetime to determine its degree of deterioration [4].

Fig. 1 Breakdown into different maintenance strategies



Maintaining an aircraft engine is not only complex and time consuming. It is, above all, expensive. Direct engine costs actually accounts for approximately 30 % of the total maintenance cost for an aircraft [5]. Maintaining a fleet of aircrafts also means challenges from a business perspective since the goals of maintenance and operations costs may conflict with desired service levels and safety levels [6, 7]. It is therefore of importance that each maintenance event is as efficient as possible to lower the costs and to be time efficient without adventuring the safety issues. On the other hand, it is also of major importance not to perform excessive work and/or component replacements and thereby throw away components with remaining life or to reduce engine availability.

2 Current Maintenance of Engines

Aircraft engine maintenance can be carried out at three separate maintenance levels [8]; the Operation level (O-level) is the lowest level activity and is carried out in the flight-line environment. For example are onboard engine performance monitoring equipment used to record engine and aircraft performance data at this level in order to detect defects or the need for routine engine maintenance [9]. At the O-level, the main focus for the maintenance is to perform scheduled and unscheduled inspections of the engine while it still is placed in the aircraft. This level also includes repairs, replacements and services which can be performed while the engine is still installed in the aircraft. The next level is the intermediate level (I-level) and the highest level is called Depot level (D-level) [10]. Main focus for the I-level is scheduled and unscheduled maintenance and to repair or perform service on line-replaceable units (LRUs) that can be performed without sending the engine or LRUs to D-level. D-level is the level were larger overhauls and maintenance of LRUs can be carried out. Also are inspections, services and replacements and repairs of shop-replaceable units (SRUs) are also performed at this level and normally D-level is additionally responsible for spare part distribution.

Aircraft engine maintenance has historically been carried out at fixed time intervals between major overhauls, but has then moved on to be carried out when needed, with no fixed time intervals [11]. Instead, services and controls of the engine system have been implemented according to a service plan to reduce the number of maintenance occasions to not perform excessive maintenance and only maintain the engine when needed.

In the aviation industry two main directions can be identified, the civil aircraft industry and the military aircraft industry. The aircraft engines used in both these specializations are based on the same techniques and constructions. The military engines are however exposed to higher loads, and thereby higher life consumptions, then the engines in the civil aviation industry. A military aircraft during a flight mission can for example vary its flight altitude many times, while a civil aircraft normally starts and climbs to a specific altitude until it descends to land.

Federal regulations govern all aircraft engine related matters. To maintain an aircraft mainly three sets of standards need to be fulfilled. First the standards in the manufacturer's Federal Aviations Administrations (FAA)-approved maintenance manuals [12]. Next are the standards for the maintainers' FAA-approved progressive inspection and maintenance program that must be met. Finally, the maintainer must meet the additional airworthiness standards from the Code of Federal Regulations (CFR) as well as the regulations concerning records, personnel and working conditions [13].

3 Selection of Maintenance Tasks

An aircraft engine consists of three different categories of components; Life Limited Parts (LLP), On-conditions Parts (OC-parts) and consumables (see Fig. 2). LLPs are components with a fixed life limit and are exchanged when they have reached their life limits [12] since they are safety critical (i.e. a breakdown may cause an engine breakdown that are so serious that it would cause an aircraft crash). OC-parts are "stochastic" parts that are approved for further use as long as their condition is within approved limits. There can also be scenarios where a LLP has not reached its life limit, but cannot be approved for continued service due to other aspects as cracks, fretting or similar. It should be noted that an LLP also can be evaluated as an OC-part. The third group of components, "consumables", is a small group of components that are exchanged each time they are removed from the engine.

In order to move from fixed maintenance intervals to maintain the engine when required, an on-condition maintenance concept must be designed to guarantee reliability. This is one of the reasons that Reliability Centered Maintenance (RCM) was developed within the aircraft industry. The RCM process is designed to focus engineering attention on component level in a formal and disciplined manner, leading logically to the formulation of a maintenance strategy plan. Benefits with RCM also include the development of high quality maintenance plans with decreased lead time and at lower cost [14].

RCM methodology is used to generate and optimize a maintenance program, including inspection requirements, that focuses on preventive maintenance on the

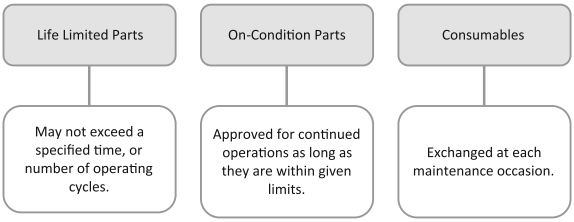

Life Limited Parts
On-Condition Parts
Consumables
May not exceed a specified time, or number of operating cycles.
Approved for continued operations as long as they are within given limits.
Exchanged at each maintenance occasion.

Fig. 2 Component categories in an aircraft engine

specific failure modes that are likely to occur. The methodology is based on the assumption that the inherent reliability of equipment is a function of the design and the built-in quality [15–19]. Theories related to RCM mean that performing maintenance not only should be performed to avoid failures, but also to prevent or at least decrease consequences caused by failures. That is why RCM focuses on retaining functions instead of focusing on the hardware itself [15, 16]. This means that RCM treats components differently depending on how important they are considered to be for the equipment and the system functions. This is also the reason why the components are divided into LLPs, OC-parts and consumables. If the probability that an event could cause large consequences for the systems, like a breakdown, components related to this event are found to have higher importance. Preventive maintenance is then used to act as a barrier to remove the consequences of failure, or at least to lower them to an acceptable level.

An implementation of RCM, The Air Transportation Association's (ATAs) Maintenance Steering Group 3rd Task Force (MSG-3) is the only process that is approved by the FAA for the development of a Maintenance Review Board Report (MRBR) for transport aircrafts. MSG-3 was originally developed for the Major Airlines, and was later also adopted by Regional Aviation Users. MSG-3 is however found to be an expensive and time-consuming process were a MSG-3 process for a propulsion system takes approximately 2000–2500 man hours. Even though this is a significant amount of time, MSG-3 has been proven to provide significant payback to operators in minimizing preventative maintenance costs [20]. MSG-3 outlines the general organization and decision process for determining the scheduled maintenance requirements initially projected for preserving the life of the aircraft, with the intent of maintain the inherent safety and reliability levels of the aircraft [21].

In order to evaluate and classify the failure modes into one of the three categories below, the decision process illustrated in Fig. 3 is used [22].

1. Safety related
2. Outage related, were the system not will fulfill all its requirements
3. Economic related

If a failure mode is found to be safety related, design modifications are mandatory. For failure modes within bullet 2 and 3 above, the maintenance options can for example be time directed tasks as on-condition based maintenance, run-to-failure, and design modifications [22].

While operation experience is accumulated, additional adjustments may be made by the operator to maintain an efficient maintenance schedule [24]. The ATA MSG-3 (2207) states that the objectives of scheduled maintenance of aircraft are [1]:

To ensure realization of the inherent safety and reliability levels of the aircraft.

- To restore safety and reliability to their inherent levels when deterioration has occurred.

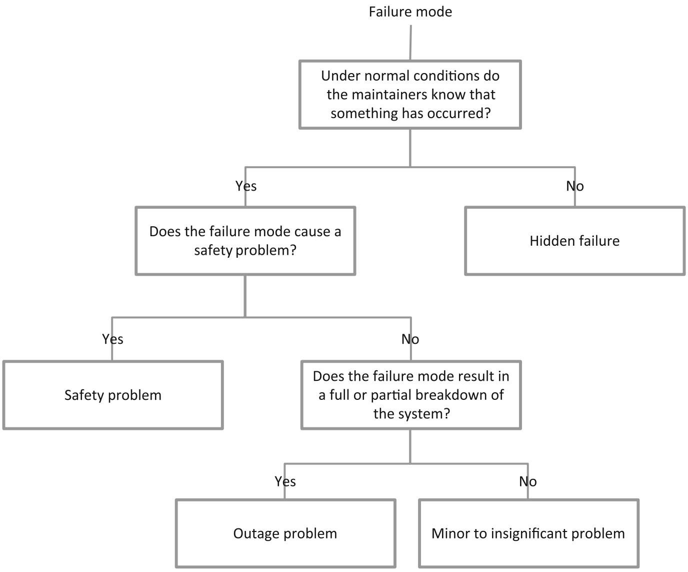

Failure mode
Under normal conditions do the maintainers know that something has occurred?
Yes
No
Does the failure mode cause a safety problem?
Hidden failure
Yes
No
Safety problem
Does the failure mode result in a full or partial breakdown of the system?
Yes
No
Outage problem
Minor to insignificant problem

Fig. 3 Decision process for a RCM program. *Source* [23]

- To obtain the information necessary for design improvement of those items whose inherent reliability proves to be inadequate.
- To accomplish these goals at a minimum total cost, including maintenance costs and the costs of resulting failures.

Finally each aircraft, and thereby also its engines, has its own maintenance requirements which are designed to keep the aircraft in an airworthy condition. These aircraft maintenance requirements typically originate from the aircrafts' manufacturer and can be revised throughout the life of the aircraft by the manufacturer, the FAA and/or the Maintenance Review Board (MRB) [2].

4 The Need for Accuracy in the Reamaining Useful Life (RUL) Prediction

The main drivers for the development of a failure prediction concept are the costs of a delay, or cancellations, of an aircraft departure or arrival. Delays can be caused by unscheduled maintenance between aircraft arrival and departure.

The purpose of failure prediction is to give the aircraft operator the opportunity to repair or replace a system during scheduled maintenance, if the system is not yet broken but are predicted to be before the next scheduled maintenance. The maintenance case is as follows:

1. A fault happens in flight.
2. Sensors detect the fault and report the fault to the cockpit.
3. The pilot/aircraft sends a maintenance request to the airport.
4. A maintenance mechanic checks the aircraft, when it is on ground.
5. The mechanic performs a fault search and a fault diagnosis.
6. Spare parts are ordered and a repair plan is made after the fault has been identified.
7. When the spare parts arrive, it is possible to carry through the repair.
8. The aircraft is ready again after the repair.

It is possible that the fault identification, diagnostics and spare parts management take too much time, so that the aircraft departure is delayed or even canceled. A cancellation or delay causes significant costs for an aircraft operator.

However the RUL prediction must match the opportunistic maintenance performed as a consequence of planned overhauls or similar actions. Indeed, when an aircraft engine is sent to D-level for overhaul, either a LLP has reached its fixed life limit or something indicates that something is wrong with the engine—in which case the engine must be taken apart, further inspected and maintained. Oil supply to critical parts, such as bearings, is vital for a safe operation. For monitoring fuel and oil status, indicators for quantity, pressure, and temperature are used. In addition to these crucial parameters, vibration is constantly monitored during engine operation to detect possible unbalance from failure of rotating parts, or loss of a blade. Any of these parameters can serve as an early indicator to prevent component damage and/or catastrophic failure, and thus help reduce the number of incidents and the cost of maintaining aircraft engines [25].

A maintenance occasion were a specific component needs to be removed makes it however, often, necessary to remove other components to be able to removed the component that needs to be maintained. This creates an opportunity to perform additional maintenance which may be beneficial in a larger perspective. Each maintenance occasion is for example related to fixed costs as leasing a spare engine, transportations, and administration. It can therefore be of interest to perform more maintenance at this specific maintenance occasion, so that this cost does not appear more often than necessary, i.e. to avoid sub-optimization by performing the right amount of maintenance at each maintenance occasion. To be able to calculate a correct maintenance schedule for what to repair, at a specific maintenance occasion, the estimated life limit for all relevant components must be available. At present is though not life estimated available for all components since only the LLPs have a fixed life limit defined, while the OC-parts instead are approved for continued operation as long as they fulfill their requirements. It would thus be beneficial, from a maintenance planning point of view, if estimates of the remaining life for the OC-parts would also be available when planning a maintenance event.

Research within this area has for example been addressed by Enright et al. [26] presented an approach for improving probabilistic life prediction estimated through the application of prediction methods. Actual F-16/100 usage data from flight data records were integrated with a probabilistic life prediction code to quantify the influence of usage on the probability of fracture for some engine component. Bolander et al. [27] on the other hand developed a method to predict the health of aircraft engine bearings, and their remaining useful lives, using spall detection.

Aircraft engines are maintained at D-level by companies specialized in aircraft engine maintenance. These companies' benefit on how much maintenance and spare parts they are able to sell. It can therefore initially be difficult to see how performing too much maintenance could be unfavorable for them. But engine maintenance relationships are built on long term basis, where both the engine operators and maintainers benefits from doing the right amount of maintenance at the right time. It is therefore of interest to both parties to perform the right amount of maintenance since the engine operators' goal is to maintain the engine with as low Life Cycle Cost (LCC) as possible without endanger the safety aspect. The maintainer, on the other hand, has an interest in performing the right amount of maintenance to ensure customer safety, but also to be able to attract new customers, make profit and to be competitive with other aircraft engine maintainers.

5 Proposed Framework

A need for better life estimates for the OC-parts has been identified and a framework on how to estimate these life predictions will therefore be developed.

Large amount of historical data of failures and replacements of components and subsystems are available since aircraft engine maintenance if strictly registered. This data could be used to provide reliability analyses and reliability predictions for the components and subsystems. This would give more accurate predictions on how much longer the OC-parts could be kept in operation before being maintained and/or replaced.

In addition, the use of physical parameters that are monitored during the operation of the aircraft engine is of interest as well as parameters that are inspected during the maintenance. Both these kind of parameters could possibly by analyzed by using Proportional Hazard Models (PHM) from the aircraft engine operation and maintenance process as covariates.

This are two separate approaches on how to better estimate the life predictions for the OC-parts in aircraft engines, and this research aims to determine which approach that is the most suitable, or if they can be combined to reach better life predictions for the OC-parts. Independently of which approach that is used, the idea is to work with a hierarchy's model, starting with an individual component up to a system level covering the maintenance process for a complete aircraft engine.

6 Conclusions

Aircraft engine maintenance can be both complex and time consuming since each aircraft is an advances system with excessive demands on safety and reliability. It is therefore important to be as effective as possible at each maintenance occasion and perform the right amount of maintenance every time.

This paper has described aircraft engine maintenance an identified a potential for improvement within the maintenance planning. A need for research within this topic has been identified to estimate the remaining life of the OC-parts so that their use can be optimized in correlation to maintenance cost. This should be done to keep the components in operation to an optimal level.

The current impression is that RAMS (Reliability, Availability, Maintainability, Safety) modeling seems to be an appropriate technique, and that this type of data eventually could increase the accuracy of the estimates of the remaining life for OC-parts.

References

1. Ahmadi A, Söderholm P, Kumar U (2010) On aircraft scheduled maintenance program development. J Qual Maint 16(3):27
2. Sinex B (2002) Aircraft maintenance tracking system. Google Patents
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. Maillart LM (2006) Maintenance policies for systems with condition monitoring and obvious failures. IIE Trans 38(6):463–475
5. Dixon MC, Force PA (2006) The maintenance costs of aging aircraft: insights from commercial aviation. RAND
6. Knotts RMH (1999) Civil aircraft maintenance and support Fault diagnosis from a business perspective. J Qual Maint Eng 5(4):335–348
7. Wu H et al (2004) Methods to reduce direct maintenance costs for commercial aircraft. Aircr Eng Aerosp Technol 76(1):4
8. Mattila V, Virtanen K, Raivio T (2003) A simulation model for aircraft maintenance in an uncertain operational environment. In: Proceedings of 17th European simulation multiconference (ESM2003), Nottingham, England
9. Nguyen PL et al (1999) Maintenance and warranty control system for aircraft. Google Patents
10. Rau C-G, Necas P, Boscoianu M (2011) Review of maintainability and maintenance optimizatio methods for aviation engineering systems. Sci Mil 2:7
11. Ackert SP (2010) Basics of aircraft maintenance programs for financiers—evaluation & insights of commercial aircraft maintenance programs
12. Aragones JK et al (2000) Method and system for planning repair of an aircraft engine. Google Patents
13. Kennet DM (1993) Did deregulation affect aircraft engine maintenance? An empirical policy analysis. Rand J Econ 24(4):542
14. Brauer DC, Brauer GD (1987) Reliability-Centered Maintenance. IEEE Trans Reliab R-36 (1):17–24
15. Nowlan SF, Heap HF, Airlines U (1978) Reliability-centered Maintenance. NTIS
16. Moubray J (1997) Reliability-centered maintenance. Industrial Press

17. Dhillon BS (2002) Engineering maintenance—a modern approach. CRC Press, USA
18. Vatn J (2007) Veien frem til "World Class Maintenance": maintenance optimisation. NTNU, Trondheim, p 208
19. Rausand M (1998) Reliability centered maintenance. Reliab Eng Syst Saf 60(2):121–132
20. Fantasia L et al (2004) System and method for defining a maintenance program. Google Patents
21. Ahmadi A (2010) Aircraft scheduled maintenance programme development: decision support methodologies and tools. Luleå tekniska universitet, p 154
22. Tsang AHC (1995) Condition-based maintenance: tools and decision making. J Qual Maint Eng 1(3):3–17
23. Smith AM (1993) Reliability-centered maintenance. McGraw-Hill, New York
24. ATA MSG-3 (2007) Operator/Manufacturer Scheduled Maintenance Development. Washington, Air Transport Association of America
25. Tumer IY, Bajwa A (1999) A survey of aircraft engine health monitoring systems. In: Proceedings 35th joint propulsion conference
26. Enright MP et al (2006) Application of probabilistic fracture mechanics to prognosis of aircraft engine components. AIAA J 44(2):311–316
27. Bolander N et al (2009) Physics-based Remaining Useful Life Prediction for Aircraft Engine Bearing Prognosis. In: Annual conference of the prognostics and health management society, p 12

In Need for Better Maintenance Cost Modelling to Support the Partnership with Manufacturing

Gary Linnéusson, Diego Galar and Mikael Wickelgren

Abstract The problem of maintenance consequential costs has to be dealt with in manufacturing and is core of this paper. The need of sustainable partnership between manufacturing and maintenance is addressed. Stuck in a best practice thinking, applying negotiation as a method based on power statements in the service level agreement, the common best possible achievable goal is put on risk. Instead, it may enforce narrow minded sub optimized thinking even though not intended so. Unfortunately, the state of origin is not straightforward business. Present maintenance cost modelling is approached, however limits to its ability to address the dynamic complexity of production flows are acknowledged. The practical problem to deal with is units put together in production flows; in which downtime in any unit may or may not result in decreased throughput depending on its set up. In this environment accounting consequential costs is a conundrum and a way forward is suggested. One major aspect in the matter is the inevitable need of shift in mind, from perspective thinking in maintenance and manufacturing respectively towards shared perspectives, nourishing an advantageous sustainable partnership.

Keywords Maintenance · Cost modelling · Consequential costs · Manufacturing · Production flows · Dynamic complexity · Sustainable partnership · Shift in mind

G. Linnéusson (✉) · D. Galar
School of Engineering Science, University of Skövde, P.O. Box 408
541 28 Skövde, Sweden
e-mail: gary.linneusson@his.se

D. Galar
Luleå University of Technology, 971 87 Luleå, Sweden
e-mail: diego.galar@ltu.se

M. Wickelgren
School of Business, University of Skövde, P.O. Box 408
541 28 Skövde, Sweden
e-mail: mikeal.wickelgren@his.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_20

1 Introduction

In the eye of financial accounting all activity in business is cost generating, maybe despite invoicing. All aspects and functions of a firm consume resources and add costs, in that sense maintenance and manufacturing are alike. However resources can be wisely or unwisely utilized, and thus add more or less cost. One problem in manufacturing though is that depending on how production flows are defined costs related to maintenance and downtime appear differently [1–4] and can be difficult to unveil due to its interconnected nature with delayed feedback. In these interconnected systems a narrow analysis utilizing "localized cost reductions" may intrude severe consequences into the entire system of production; thus a thorough analysis judged upon ability to generate throughput must be performed before any cost reduction is put into effect [5]. It underlines the need of dealing with the wider perspective of maintenance costs than "localized cost reductions" including the systems perspective.

Manufacturing and maintenance are tightly interlinked, meaning that actions at one partner brings consequences to the other. However, often manufacturing and maintenance are managed with separated budgets, making it hard to identify necessary jointly optimal procedures for a win-win situation. Manufacturing is the partner in the relationship that adds value to the product and maintenance is its supporting function ensuring the required capability of the machines. Despite how well or poor maintenance is carried out it costs money, and on the individual activity level it is difficult to justify and thus often seen as a cost function only [6]. It brings the consequences that minimizing maintenance cost focus direct costs and neglects the more tangible part of costs; consequential costs from minimized maintenance interventions due to minimized budgets.

Literature points out the importance to convert maintenance costs, but nonetheless values, into cash terms in order to support the communication on the language of higher management, which is money [7–10]. In order to value maintenance correctly its long term effects on the interaction in the organization must be included in the evaluation. However, managing a company focus utterly on how to maximize ROI (return on investment) and cash flow. And, the reductionist approach of the traditional financial control has brought a too short sighted focus on cost reduction instead of the organizations long term survival [11, 12]. It comes down on maintenance with bad consequences; supported by cost over profit maintenance performance measurement systems that do not capture the reverberations of today's actions [9]. That saving money today is so easy, on the behalf of delayed impact effects suffering much more expense later, is unfortunately not acknowledged by most writers, according to [9]. Furthermore, it is considered very hard to identify the contribution to company profits from the maintenance budget on the macro level [6] and in combination with the short sighted cost reduction focus it may be what hinders identification of well-functioning strategies. Thus in need to be illuminated further.

An example on the development of well-functioning maintenance system is described in [13] also published in [14], showing the need of a shift in mind at problem stakeholders, i.e. the organization facing the problem phenomena of a poor maintenance system, in order to bring about desired development [14]. led a modelling endeavour at a large company and the process brought several important conceptual shifts in the way they viewed maintenance and thus their focus in the study went from cost minimization to centre on the physics of breakdowns through equipment degradation [13]. On that theme [15] states that; in order to strategically manage cost optimization of the effects of maintenance on equipment, machine or even company level we must adhere to the fundamental multi-disciplinary ingredients of: understanding the mechanisms of degradation, quantitative models that concern impact from actions such as operations and maintenance have on degradation, and strategic management of maintenance. This paper acknowledges the importance of these mechanisms and the value of dress them in economic terms in order to improve visibility of consequential costs for higher management. However, it is a challenge to define those mechanisms in small enough rates of precision in order to build trust into such model's results, previously tackled by including stakeholders in the modelling process [16]. That is why first steps identified have to be towards better maintenance cost modelling and not an absolute panacea equation solving the problem, most likely such method will never be developed.

The foundation of this paper is thus based on the beliefs that maintenance efforts must be valued through the lens of both direct cost and consequential cost with a life cycle perspective in the context of the organization it interacts. Preparation of maintenance efforts start perhaps in the acquisition process and are ongoing in the daily operational business, in which decisions upon improvement on how to meet key performance indicators such as mean waiting time, mean time to repair, and mean time between failure etc. are taken. Constantly the status of equipment assets are changing and constantly the maintenance organization has to deal with it; and there are better or worse strategies depending on how the production is configured. However on a general level it can be stated that the better the status of machines and procedures the less emergency work and breakdowns, and reverse the worse status of machines and procedures the more time spent on emergency work and breakdowns [13, 14].

This paper acknowledges that it is the decision making made by humans that is key leverage in the development of manufacturing systems [17] and thus the performance of maintenance, so deeply involved in it. Another important aspect when it comes to practical implementation and change, considered important to acknowledge, is to adhere to the need of building organizational capability that in turn will lead to sustainable high performance [18], also in line with Lean practices empowering the front line teams [19]. However, the slow process of building capability bottom-up must be combined with top-down, results-driven change, as according to [20]. It motivates the need of visualizing maintenance total costs on short and long term (organizational capability building) for decision makers, and underlines its potential to bring mature strategic thinking on how to confront current situation for modern manufacturers.

Maintenance is a support to production in order to facilitate maximum usage in manufacturing, however at all stances it is not always necessary to require most possible reliability because service level agreements tell so; it is a matter of cost and total performance of systems that finally count in the financial statements and it should be the criteria also when defining these systems in practice. The relationship between manufacturing and maintenance has to be re-evaluated in the view of both sides of the partnership in order to identify more cost effective procedures.

This paper begins on showing the prominent significance maintenance plays in order to stay competitive on the global market as an underlying background to the conceptual ideas presented later on the potential of visualizing maintenance total costs for higher management for better decisions on maintenance strategies.

1.1 What Modern Manufacturers Confront

The situation for actors within manufacturing industry is strained, with the combination of escalating competition on a global market and growing demands on return on investment. In short the constant pressure of making more with less is constantly present. New advanced technology and automation has during the latter years been the means manufacturing companies have applied in order to manage and stay on market. With its development production throughput has been increased, both by combining multiple tasks previously carried out in a functional layout and through improved production flows. It has contributed to the reduction of inventory and stock levels, reduction of man hours in production and increased quality and precision; thus delivering more with less. However, in order to sustain on the global market further efforts will continuously be needed in order to stand competition. Furthermore, the development until today has dramatically changed the arena for how to play the game in maintenance. In which the flow of new advanced technology and increased automation brings increased demand on competence for the maintenance personnel, and the failure rate patterns are more complex to understand and increased requirements on equipment- and component-reliability are prominent. The maintenance function is also challenged from the perspective of decreased time buffers in production systems, in order to maximize throughput and cash flow, adding increased vulnerability when a breakdown take place. In combination these aspects bring forward requirements on new levels of perfection on maintenance, both technically and organizationally. It definitely points out the importance of maintenance management as an enabler for future contribution to the improvement of the input/output ratio. There are numerous strategies for how to run production and how to include maintenance in order to encounter the optimum ratio, and how these strategies are valued strongly decide upon choices for future action. Viewing manufacturing and maintenance as separate negotiating parties may hinder systems' full potential, also called sub-optimization, and makes it harder to value benefits on the level of totality. However, maintenance management is a complex undertaking with the inherent difficulty of long delays between cause and effect in which a decision apparently beneficial on short term may end up in costly and repeating actions during the life-time of an equipment. The efforts spent should thus constantly be valued by its economic life cycle effects. It puts focus on the ability to evaluate maintenance strategy and its consequences over time in order to convert it to financial statements in a convincing manner for decision makers.

1.2 A Management Decision Making Perspective

Maintenance holds many criteria for being hard to manage on the operational level: stochastic behaviour in the deterioration and failure process of equipment, large portion of unplanned events, a sudden failure in critical equipment can make any important long-planned activity delayed, decisions of importance may have to be taken quickly with no time of thought, management is under constant time pressure, thus little time for abstract and strategic thinking [6].

Another aspect is the phenomena of worse-before-better dynamics [21], common in complex systems also present in maintenance systems. If management don't understand why it occurs and for how long the short-run "worse" deterioration of system performance might last, it may instead be held as evidence on that the new strategies don't work and are abandoned [13]. Our human ability to acknowledge system dynamics is strongly limited and we construct mental models of phenomena in order to cooperate and understand the environment we are part of [22]. However, the mental model of two persons observing the same aspect may differ in their description of it because they looked at different things [23]. Fortunately, if it is put into light that decisions are being based on these incomplete, partially shared mental models with lack of coherence and comprehensiveness they can also be subject to learning and improve decision making around maintenance [14].

Decision making in maintenance includes many stakeholders, from operators and machines producing value to investors interested in maximized profit via the board of directors, top management, and operational management and support, see illustration in Fig. 1. At each level decisions are taken, however their effect may vastly vary. The lower triangular in the figure is most often regarded in everyday work, however the upper triangular represented by stakeholders interested in the company development are also actively part of daily agenda however more hidden. All with different roles and with different incentives for being part of the structure. It is the operators and machines that utterly create the value orchestrated by the requirements from a level above, ending in the interest of for instance pension savers on ROI.

The Fig. 1 serves as a basis for visualizing structures having an effect on the decision making in different levels, and the incentive structure's influence on thinking in respect to possible short term effects and the organization's long term thinking capabilities. It may be further exemplified by the following simplified story in order to bring sense into seemingly wrong decisions, depending on what you

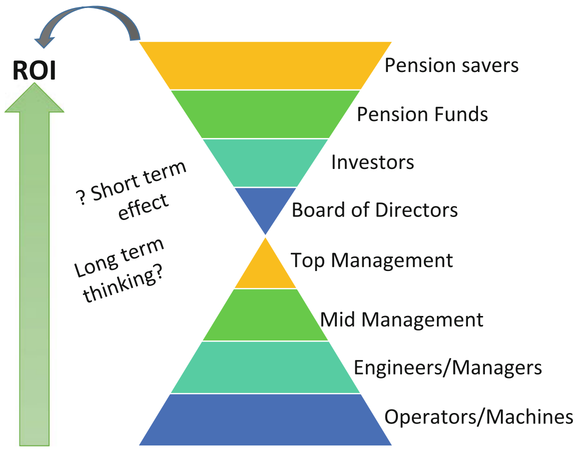


Fig. 1 Illustration of incentive structure

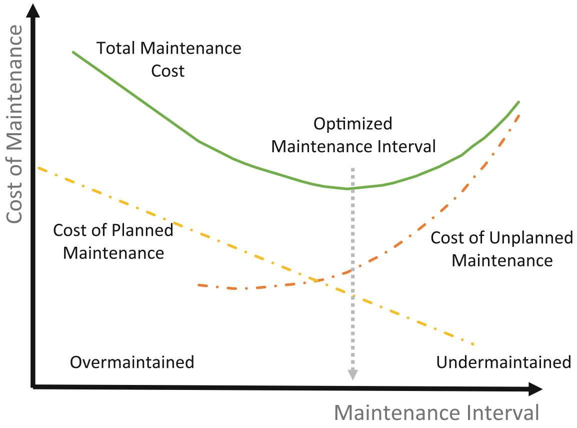
choose to see (mental models). The scenario plays in a maintenance organization of 300 persons. A strong Investor in the company points: Why do they need to be so many persons in maintenance? They cost a lot. The Board of Directors do nothing more than repeats the question to organization. Top Management ask their people: Why do we need so many? In order to show vigorous command: Cut 1/3. On short term nice performance measures are presented, and most importantly, better cash flow. No explicit consequences in production are noticed. The measures in financial control-terms confirms that it was a correct decision, our suspicions were right. Eventually problems add up and number of breakdowns increases. Workload shifts sneaking from preventive towards unplanned maintenance, and cost per piece increases due to less throughput, poorer quality, etc. Case is highlighted at top management, decision maker using financial control as performance measure observes this negative trend and applies present mental model that worked last time, commanding: Cut more personnel in maintenance. The vicious circle of decreasing maintenance capabilities to perform accelerates. A simplified story near all maintenance people have experienced in small or in large. A natural development in a context of complex systems in combination with applying your own mental models of reality without thorough analysis so important, perhaps due to lack of time and tools for better analysis.

2 Maintenance Cost Modelling

2.1 Maintenance Consequential Cost

Since financial accounting started over hundred years ago [12] direct maintenance cost have been explicit to account and is part of standard accounting procedures [3]. However maintenance that is not optimally performed also cause indirect costs in

Fig. 2 Cost minimization graph, based on [19]



the organization, termed consequential costs [4], and these appear in other parts of the organization such as for instance more time spent by production man hour, less produced products or poor quality. A graph of the trade-off between direct and indirect costs of maintenance can in theory be illustrated as in Fig. 2 below.

In an optimal procedure the direct costs, represented by *cost of planned maintenance,* should be balanced against the consequential costs represented by *cost of unplanned maintenance* [19]. However in practice identifying this trade-off is not straightforward. Consequential costs generated from poor or absent maintenance is implicit and intangible to its character, difficult or near impossible to track in an accounting manner. The need of better estimated quantification of consequential costs in order to support manufacturing with optimally performed maintenance was defined long time ago, however a problem of inherent subjectivity [4]. Maintenance consequential costs are suggested into four categories [4]:

1. Associated resource impact costs, productivity loss (loss in production time) at machines/equipment that are connected with the resource in which the failure has occurred.
2. Lack of readiness costs, machines/equipment that for the moment are not used, thus in idle, but not in a ready condition for production—seen as an incitement for keeping capital investments in shape.
3. Service level impact costs, occur when replaceable resources/equipment in a pool of resources fail and those left in operation must be utilized in a more costly manner due to that they are fewer and the work still must be done to maintain the required level of service.
4. Alternative method impact costs, in times of great pressure for delivery failure in one machine/equipment may force usage of alternative methods in order to deliver on time, however it may be performed in a non-optimal procedure.

How these consequential costs can be calculated is also suggested and categories 1, 2, and 4 are represented by time dependent impact profiles in which estimations of “time from failure to start of impact” and approximated cost accumulations during the impact period have to be applied to the certain case [4]. Category 3 is represented by an equation and an optimization problem in order to balance the temporary increased loads on the equipment due to failure in any resource and at the same time maintain the service level required. Traditional cost models has a linear approach and assume the cost of lost production to dominate the downtime costs, neglecting consequences such as safety buffers, wasted raw material due to scrap, etc. [3]. The non-linear description of consequential costs by [4] is a step towards a more practical approach attempting to bring a better estimation.

Plainly looking at definition of maintenance cost is scares and has resulted in more qualitative designations in literature dividing maintenance costs into two general categories [3, 8, 24]: *direct costs* also known as *intervention costs* (maintenance operations including labour, administration, material, subcontracting, to name a few); and *downtime costs* representing more or less all consequential costs (production losses, reduced quality, etc.). There are example of listing costs into much more detail as well [25]. Instead it seem, literature in maintenance cost and modelling refer to these two general categories and states what kind of considerations regarding cost or delimitations on what to include constitutes each model [7, 8, 10, 24, 26].

2.2 On Maintenance Optimization Models

This section will provide an overview of difficulties worth considering regarding maintenance optimization models. A maintenance optimization model is considered a mathematical representation including quantified maintenance costs and benefits in which the optimal balance is found [6]. However, there is also an emergence of applying simulation for maintenance cost modelling [1]. And it is also argued for the increasing need of bringing optimization into the system of maintenance management [9].

In his review of maintenance optimization models [6] identifies several gaps between theory and practice worth to bring up:

- It is difficult, even for technicians and managers, to understand and interpret the stochastic nature of maintenance optimization models.
- Little attention is paid to make models understandable to practitioners, instead focusing on mathematical analysis and techniques rather than solutions to real problems.
- Too few industrial problems are addressed by academics due to low incentives addressing real problems and that companies are not interested in publication.
- There is a lack of knowledge on which models are suitable for which practical problems, attention should be on collecting present knowledge into one model and review its applicability rather than publish new ones.

- The cost for doing an optimization analysis is not always worthwhile compared to its savings, and often results from optimization show on lowering indirect costs not beneficial to maintenance.
- Optimization models often involve wrong kind of maintenance (planned revisions and overhauls) which have not always proved effective. Better designs and condition-based maintenance has resolved parts of the problem, however dependent on better prediction capacity.

As it seem, history of maintenance optimization models have struggled with practical applicability. In recent time the perspective of quantitative modelling has also been claimed being extremely specific [10]. Further, other recent conclusions on maintenance optimization models include that they are often over simplified in the aspect of: using fixed value for cost of corrective and preventive maintenance, and excluding time spent on repairs [24]. Also, the lack in literature of including degradation patterns and the effects of system improvement to include a more practical and real behaviour concerning aspects of asset management [24].

In their paper [1] present a large review over applications of simulation in maintenance research and conclude on the potential of discrete-event simulation (DES) to address cost reduction. The dynamic capabilities of DES make it possible to compare different strategies including several aspects with the ability to evaluate systems in an integrated way, such as impact of condition monitoring on staffing. Their conclusion is that DES should be able to bring insight into maintenance systems on the level that it has brought into manufacturing systems.

In short the research front in the area of cost modelling using DES for manufacturing companies today combines: production engineering data, financial data and optimisation technology with innovation in order to bring about useful information and knowledge facilitating better informed decision making within the development of production systems [21].

2.3 Value of Maintenance

Maintenance is often claimed as a value contributor to a company and research on showing its value instead of the focus on cost is common. This is performed in different ways, cost models to value financial impact in planning improvements in maintenance [7], maintenance performance improvement [8], estimate added value from maintenance services [27], and etc.

Another aspect in order to value maintenance, or the reliability of an investment, through the lens of direct and consequential cost in a life cycle perspective is the using LCC (Life Cycle Cost). However, in their review of published case studies of life cycle costing [28] it shows how difficult it is to make sound LCC studies. Purpose with LCC is to support the more correct long term cost of ownership [29] calculating the investment cost but also net present value of running cost, illustrated

Fig. 3 Illustration of costs from an investment

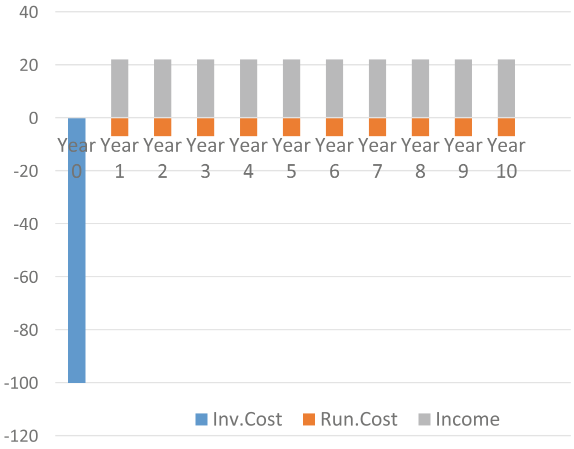
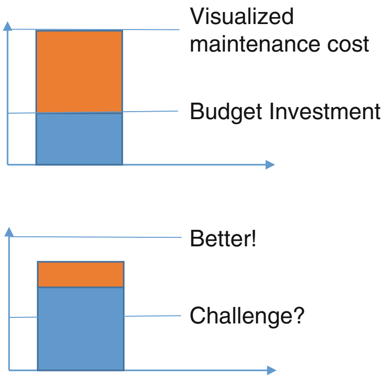


Fig. 4 Illustration of LCC scenario



in Fig. 3. It is preferably applied in the early acquisition phase in order to take better informed decisions on specifications on reliability of investments.

This can be further illustrated by Fig. 4 that shows an example scenario: the upper graph shows investment costs and estimated maintenance costs on budget, the lower graph shows the same investment after have applied LCC illuminating on the benefits of increased reliability on the behalf of a larger initial investment.

Besides tools like LCC maintenance concepts have emerged that explains the function and value of maintenance, such as: TPM (Total Productive Maintenance) [30], Lean Maintenance [19], Value driven maintenance [31], and Asset management [32]. These concepts include aspects of practice, and management in order to get theory into practice, thus important descriptions of thought for any company interested in developing their maintenance function. Yet, it is not always straightforward to commit resources in order to implement these concepts and systems in the organization, it is a matter of persuasion and commitment along the way putting focus on the requirement on understanding and ability to interpret social systems [22].

3 Feedback Thinking

Due to the lack of feedback thinking in the cost over profit maintenance performance measurement systems the importance of it is included in this section. A qualitative CLD (Causal Loop Diagramming) model by [33] is used as base for exemplifying feedback thinking and its relevance to dynamic analysis in maintenance. Their model is based on numerous case studies on the core issue of organizations' ability to implement improvement programs such as for example TQM (Total Quality Management) into their everyday activities. Their research shows that key for successful implementation is independent from choice of improvement tool, identifying the root cause being a systemic problem requiring much effort in managing the interaction of tools, equipment, workers, and managers. Thus the model shows considerations possible on a generic level to draw conclusions from into the aspects of implementing a better maintenance function, meaning it is applicable to exchange TQM into any other of maintenance concepts mentioned earlier as well. The model will be described in brief and if deeper understanding is wanted it is recommended to read the paper of [33].

The underlying structure in the model is the physics of improvement, seen in Fig. 5 in which an asset's capability is the fundamental base for performance. It is the *Capability* of an asset, such as process, machine, working procedure, etc., together with the *Time Spent Working* that result in the *Actual Performance* from that asset. *Capability* is represented as a stock (rectangle mean stock or level carrying the current state of the system [34]) that accumulates improvements over time in relation to the *Time Spent on Improvement*. Either *Actual Performance* is increased temporarily through more working hours, or efforts are invested in improvement work which brings a more lasting value to the asset through increased *Capability*, thus with gains to production capacity with longer duration. However, *Time Spent on Improvement* are not immediately equivalent to increased *Capability*, it takes time to eliminate root causes or to structure a preparation of preventive maintenance and it takes time until effects from that improved work is gained;

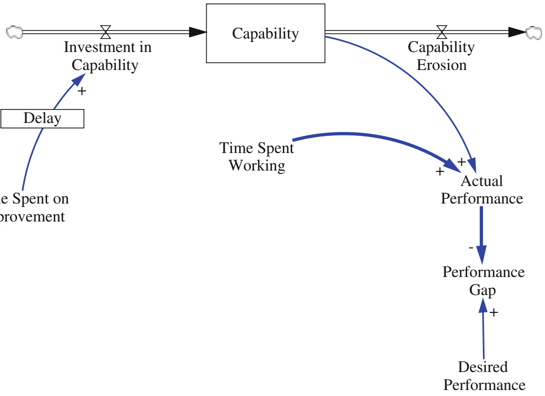


Fig. 5 The physics of improvement, based on [33]

illustrated by the *Delay* in the model. Furthermore, *Capability* of an asset is under the law of deterioration, machines wear, working procedures become obsolete, etc. showed by the flow of *Capability Erosion*. Depending on technical and organizational complexity the delays vary, less complex process improvements having shorter delays of few months, while high complex processes such as product development may carry delays of several years. Anyhow, the connection with maintenance are close, for instance: time spent on improvement identifying root causes to breakdowns and its countermeasures, or systematically updating current base conditions and initiating measures of preventive character are put in place with the aim to balance the resources towards planned work in order to interact with the customer's assets more smoothly avoiding disturbance in production. And, it takes time until the harvest from such work can be gained. However, these actions serve the common goal of delivering *Desired Performance* set by senior managers, and the difference between desired and actual is the *Performance Gap*. In maintenance these gaps can consist of backlogs of preventive maintenance, corrective maintenance, breakdown analysis, requirement specifications for new acquisitions, etc.

The Fig. 6 represents the model in its totality, including three balancing loops (B1–B3) and a reinforcing loop (R1), and some three more variables. A balancing loop in its character includes goal-oriented dynamics that strives for closing a gap. A reinforcing loop in its character accelerates growth or decline depending on its direction.

B1, any work senior management want performed create a *Performance Gap* followed by increased *Pressure to Do Work* in order to close the Gap. Thus increased pressure intensify reporting performance measures on for instance output of individual machines sending explicit signals of what is important. Enforcing action and *Time Spent Working*, increasing *Actual Performance* in order to close the

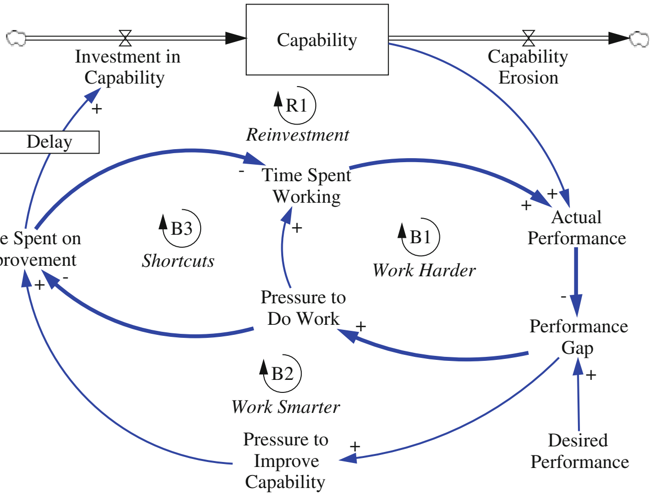


Fig. 6 The capability trap model, based on [33]

Performance Gap. The balancing loop of *Work Harder* strives towards the goal of closing the gap through increased pressure on short term results.

B2, the *Performance Gap* also initiate actions to improve assets capability through increased *Pressure to Improve Capability*, which can be initiated by activating improvement programs, training or more simple actions to cut time required for performing *Actual Performance*. The balancing loop of *Work Smarter* also strives towards the goal of B1, however, as brought up earlier this action is delayed and it takes time until effect on *Capability* is generating the improved performance. The actions invested in capability improvement though are of much more enduring character and bring value in long time, similar to a well-functioning preventive maintenance system.

B1 and B2 together quit on their own explain well the dynamics of working harder versus smarter, the shorter deadlines on delivery of performance the stronger incentive for solving pressing problems by quick fixes. Surprisingly [33] discovered that, at those companies studied, working harder was no occasional lifeline used when in pressure in between peeks but rather the standard operating procedure.

R1, the reinforcing positive feedback loop of *Reinvestment* connects the *Work Harder* and *Work Smarter* loops with the arrow between *Pressure to Do Work* and *Time Spent on Improvement*. This connection is due to that organizations rarely have excess resources. It means that under continued pressure it eventually forces workers to reduce their time spent on improvement in order to manage the situation. The *Reinvestment* loop represents that if one option is chosen on the behalf of the other it is likely reinforced and develop into a permanent state; if *Capability* is successfully improved through continues improvements, the capacity to deliver *Actual Performance* will increase and freeing more time for even better improvements. As in line with continuous improvements in lean philosophy [35]. Although the other way around is also true; that using working harder efforts with less time spent on improvements the erosion of capabilities will drain the current asset's *Capability* to deliver *Actual Performance* increasing further *Pressure to Do Work*. This is the vicious cycle many get caught in but at least mentally never want to be in. It is a state in which fixing the worries of the day overload continuous action, on short term neglecting time for improvements, predictive maintenance, quality improvements, instead it results in a manner of firefighting and increasing safety stocks and buffers, piles of reports on quality errands, etc. not surprisingly failing any improvement program started. Unfortunately it is tempting to abandon the reinvestment of time in a work smarter behaviour leading to further enduring savings. Instead, resulting in new amplified requirements on *Desired Performance* as indications on that the working smarter pays off, jeopardizing the virtuous cycle of building capability into our systems.

B3, represents the final explanation to the behaviour of why most organizations go into the capability trap [33, 36], and it is the balancing loop of *Shortcuts*. Due to that *Capability* does not drop immediately and in combination with increasing pressures small shortcuts are taken to win time from improvements towards working; skipping an improvement team meeting here, produce on a maintenance scheduled window in production there, eventually cutting all corners there is to cooperate with the development of pressures and lost capabilities. This interconnection is represented in Fig. 6 with a negative link enabling closing the *Performance Gap* in yet a third procedure. The shortcut loop is effective and tempting to use in order to close the throughput gap, on short time there are numerous aspects that if skipped gain time without any harsh consequences due to the momentum in developed capabilities during the path of history.

The capability trap represented by the shortcut balancing loop, in Fig. 6, is a behavior carried out in many stances at the same time in an organization. It could be questioned if development of this behavior isn't seen by managers, unfortunately they do not realize how deeply it has trapped them and use countermeasures reinforcing it even further [33]. Several things interact in these dynamics and some are due to human behavior, that people:

- generally falsely assume cause and effect to be closely related in time and space [13, 33]
- often look for explanations for a puzzling event in nearby recent events assumed to have triggered it rather than underlying patterns of behaviour [14]
- simplify and tend to assume a single cause from each event [33]
- underestimate the effect of time delays [14, 33]
- and omits feedback processes in actions [14, 33, 36]

In total, it leads to self-conforming attribution errors blaming the people instead of acknowledging the failing dynamics of systems resulting in degradation of process performance [36], which is required in order to define accurate measures for capability development. It makes it difficult to analyse the system of study in the manner of identifying root causes behind the behaviour. Therefore, the authors [33] argue the need of a shift in mind in how to manage the capability trap is required; in order to acknowledge the underlying structure of behaviour, and implement improvement efforts on the conditions of these structures.

Considering these aspects system dynamics feedback thinking can visualize the system of study, and as previously stated by [37], we can answer questions of the character: -what causes our recurring pattern of behaviour? There are some examples of approaches of this character to maintenance issues [13, 14, 27, 33, 37, 38] with learning perspectives. Which is considered important in order to bring about correct countermeasures in real systems.

4 Industrial Problems to Address in Future Research

The industrial problems to address in this research can be represented by:

1. Visibility of consequential costs
2. Deal with the inherent problem of delay between action and its consequence effects in the system of study
3. Partnership between manufacturing and maintenance (and financials)

Visibility of consequential costs, there is a need to make important capabilities in maintenance visual for decision makers in forming their policies and strategies for the company. The financial statements (balance and income) lack in order to address accurate maintenance strategies from the information they provide. Unfortunately important assets are not possible to account in the financial statements [12], such as:

- Working procedures are not visual and not valued
- Knowledge is not visual and not valued
- Machine status are not visual and not valued
- Value of databases is not visual and not valued
- Ratio of planned/reactive maintenance is not visual and not valued
- Policies and decisions that concern maintenance ability to perform its mission are not visual and not valued

Thus, many capabilities in the organization belonging to assets in maintenance are not valued and hidden for decision makers in higher management. However, it is relevant information when deciding upon budget conditions for maintenance. As learned from the Capability Trap Model, in Fig. 6, the balancing of resources in short term or long term actions have severe effects on the development of a system's capability. System dynamics models generally serve the purpose of learning about the dynamic complexity of system in study, however examples of research also aiming on visualizing added value of maintenance services [27] or valuing machine strategies [16]. Both perspectives are regarded useful in order to bring about strategies based on underlying system behaviour and at the same time communicate on the condition of decision makers, which is money. Applying simulation allow also a life cycle cost perspective with the capability of assets in focus, how these are maintained and their current condition result in behaviour [13].

Deal with the inherent problem of delay between action and its consequence effects in the system of study, the maintenance function is comprehensive and include many interacting aspects with manufacturing. The instrument of financial statements act short on connecting short term cash flow improvements with its consequences, thus complementing methods will support. In order to construct alignment between maintenance strategies and their financial impact following remarks illuminates on the need:

- The practical fact of delayed consequences in the maintenance function requires a method able to include those delays
- Feedback from consequences must be acknowledged in order to correctly value decisions regarding the development of maintenance
- We must approach peoples mental models of the situation in order to improve the connection between financial statements and maintenance

We use our mental models in interaction with each other and the systems we act in, and decide upon. Each individual act based on her own simplified model of the system of study (reality) and ability to understand from her perspective and available information at that time [22, 23]. In combination with our human disabilities to mentally simulate complex systems the importance of visualizing them are even more highlited.

Partnership between manufacturing and maintenance (and financials), a more systemic view upon the two partners hold potential of avoiding localized cost reductions [5]. There is an ongoing tug of war between manufacturing, maintenance, (and financials) in respect of time of equipment intervention, strategy of resources and budget. And, at the same time a need of a mutual relationship in order to value action on a systemic perspective. It can be argued that TPM contribute to improved partnership, and its attempt to integrate and align maintenance and production goals act in such direction. However the partners of manufacturing and maintenance are managed from their own budgets in specifics, which tend to limit their actions to focus on their own perspectives. Therefore, some general thoughts on aspects to regard in order to reveal further potential from improved partnership between these two partners:

- Deeper understanding of one another's need is required in order to attain maximum output (ROI)
- How the partnership between manufacturing and maintenance is set up directly, and indirectly, define its results
- There is a need to facilitate the analysis of manufacturing and maintenance as one system
- Manufacturing should be able to define their request on availability depending on current production flow in study, eliminating unbalanced maintenance need

These potentials from an improved partnership expose efforts maybe on two levels, on operational and a strategic. An identified tool in use at the industrial partners of this research that regards the value of maintenance through the lens of both direct cost and consequential cost in a life cycle perspective is the usage of LCC, however limited to the acquisition phase. LCC brings into the picture the joint costs of manufacturing and maintenance and may support a partnership in regard to trade-offs in initial definition of equipment from a cost perspective. In the acquisition phase it challenges habits such as strong efforts from financials department on minimizing initial costs and bring into the analysis, open for all partners to view, the total estimated costs from running the equipment as well. However the support on further learning and analysis according to the bullet list above can be further developed. Firstly, there is a need and potential of including manufacturing cost and maintenance cost in an evaluation tool of ongoing operations. Secondly, there is a need of an evaluation tool in order to value the implications from different maintenance strategies in respect to the totality of production cost. These two needs can be approached in several procedures, a way forward can be to further define the needs with respect to the problem and thus find a path.

Manufacturing development has good history records in the operations management area in applying DES as dynamic technique for optimizing throughput, and the development has reached multi-objective trade-off analysis with regard to cost savings [21]. Maintenance development still have a path left on that direction, considered a potential to identify [1]. Studies on the perspective of analyzing optimal maintenance in relation to cost are few [1] but emergence is on the operational level of incremental steps towards better productivity from present condition as in [39]. The dynamic complexity generated in only one production line, with numerous different possible downtime states, is vast. Thus it is difficult to use the consequential cost calculations by [4], in order to sort out the consequential costs from insufficient maintenance when downtime in any unit may or may not result in decreased throughput depending on its set up. With the number of production lines complexity grow exponentially and the need for more accurate cost modelling is prominent. Practically, the service level agreement with manufacturing decides upon the level of required maintenance in order to keep production running on an adequate level. However, without a tool able to address the dynamic complexity of possible states such negotiations risk to result in requirements based on the position of negotiating partners [2]. It places requirements on the client's competence to understand both the complexity of possible production line dynamics and the process of maintenance that deliver the support. And, in lack of such comprehensive knowledge it is taken care of through overcapacity in the system already in the design, resulting in the need of over-maintenance from start. Instead, applying DES may support on defining the appropriate individual service levels in the production flow, and bringing in the dimension of cost would make it possible to analyse different optimum suggestions on the use of redundancy, highly reliable equipment, preventive maintenance, or condition based maintenance, to name a few. However, even though DES acknowledges the dynamic environment of production systems, it is still lacking on the integration of how the dynamics of the maintenance function interact, that is including maintenance's purpose to keep assets in their appropriate state in order to perform the required function at minimum cost.

However, the tool of system dynamics and feedback thinking has shown capable on including the integration of how the dynamics of the maintenance function interact with production and manufacturing, although not on the same level of detail as DES thus the tools are likely serving different purposes. Feedback thinking can include maintenance's purpose to keep assets in their appropriate state in order to perform the required function at minimum cost. Another aspect is the dissemination of thought and learning in the organization from analyzing a sustainable maintenance strategy; that is a strength of feedback thinking.

5 Discussion and Conclusion

In order to motivate efforts and budget for cost optimized maintenance, the consequential costs of maintenance must be considered. However, it is also worth to consider how this is best approached. Is maximum benefit gained from identifying the cost of underperformed maintenance, or the cost of under dimensioned equipment? Would it not be even better to understand why the situation of underperforming maintenance is and how to improve in future and change our everyday habits and procedures towards investing in our asset's capabilities instead? However, it requires a shift in mind on how to view upon cost in maintenance and an important factor to consider is the practical implication of changed strategies in order to attain it. How is the problem solved most effective in practice? One aspect, for sure, is the aspect of how time is divided between working harder, gaining immediate positive effects, and working smarter, sanctioning investment of time on improvements that improve performance endurance. In that sense we have to dig deeper into the minds of decision makers and understand their uncertainties about maintenance strategy and its consequences. Thus:

- We need to dig deeper into the matter of visualizing maintenance consequential costs for higher management and investors in order to facilitate better sustainable strategies for action
- Short term (my budget) thinking must be replaced with a sustainable partnership applying long term development of the dynamics of the maintenance function in order to bring lasting value from maintenance and manufacturing as a joint system

In conclusion, DES offers a tool on the operational level that can be applied for optimizing production flows also from a maintenance perspective including aspects of choice in respect to cost and throughput, such as: redundancy, highly reliable equipment, maintainability, service level agreement, run to failure, condition based maintenance, and more. It would be a delightful tool for engineering management and maintenance. However it will not appropriately address the need of well-functioning maintenance strategy in the organization. System dynamics feedback thinking on the other hand offers a tool on the strategic level that can be applied for evaluating the interaction between market demand, production rate, use of resources on different actions preventive or reactive in character and their impact on system and asset capabilities. Thus such a strategic perspective could bring learning to higher management though visualizing the connection between the dynamics of maintenance, how it may be operated and developed for better performance, and its combined consequence to the financial statements. With increased learning insight on the devastating effects from short term improvements in cash flow can be attained, and future paths for investment of time spent on improvement can be substantiated. Such a tool will in the long run build confidence into specific developing actions of the maintenance function generating specific needs such as DES for optimizing the service level agreements.

References

1. Alabdulkarim AA, Ball PD, Tiwari A (2013) Applications of simulation in maintenance research. World J Modell Simul 9(1):14–37
2. Keld J, Iwar U (2001) Negotiating partnerships: increase profits and reduce risks. Pearson Education Limited, Harlow

3. Pascual R, Meruane V, Rey PA (2008) On the effect of downtime costs and budget constraint on preventive and replacement policies. Reliab Eng Syst Saf 93(1):144–151
4. Vorster MC, De La Garza JM (1990) Consequential equipment costs associated with lack of availability and downtime. J Constr Eng Manage 116(4):656–669
5. Bragg SM (2010) Cost reduction analysis: tools and strategies. Wiley, New Jersey
6. Dekker R (1996) Applications of maintenance optimization models: a review and analysis. Reliab Eng Syst Saf 51(3):229–240
7. 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 107(1):260–273
8. Salonen A, Deleryd M (2011) Cost of poor maintenance. J Qual Maint Eng 17(1):63–73
9. Sherwin D (2000) A review of overall models for maintenance management. J Qual Maint Eng 6(3):138–164
10. Sinkkonen T, Marttonen S, Tynninen L, Kärri T (2013) Modelling costs in maintenance networks. J Qual Maint Eng 19(3):330–344
11. Bengtsson L, Lind J (2013) Innovation eller kvartalskapitalism?. Utmaningar för global svensk produktion, Liber, Stockholm
12. Johansson U, Skoog M (2007) Verksamhetsstyrning—för utveckling, förbättring och förändring. Liber, Malmö
13. Sterman J (2000) Business dynamics: systems thinking and modeling for a complex world. Irwin McGraw-Hill, Boston
14. Carroll JS, Sterman J, Marcus AA (1998) Playing the maintenance game: how mental models drive organization decisions. In: Stern R, Halpern J (eds) Debating rationality: nonrational aspects of organizational decision making. Cornell University ILR Press, Ithaca
15. Murthy DNP, Atrens A, Eccleston JA (2002) Strategic maintenance management. J Qual Maint Eng 8(4):287–305
16. Linnéusson G, Aslam T (2014) Machine strategy evaluation using group model building in system dynamics. In: Proceedings of international conference of the system dynamics society
17. Linnéusson G, Jägstam M, Kinnander A (2009) Bridging a methodological gap in using system dynamics in manufacturing. In: Proceedings of the international Swedish production symposium
18. Linnéusson G, Jägstam M (2008) On applying a systems approach to manage operative improvements in manufacturing SMEs. In: Proceedings of international congress of cybernetics and systems of WOSC
19. Smith R, Hawkins B (2004) Lean Maintenance. eTextbook edn. Elsevier Science
20. Beer M (2001) How to develop an organization capable of sustained high performance: embrace the drive for results-capability development paradox. Org Dyn 29(4):233–247
21. Pehrsson L, Ng AHC, Stockton D (2013) Industrial cost modelling and multi-objective optimisation for decision support in production systems development. Comput Ind Eng 66 (4):1036–1048
22. Forrester J (1971) Counterintuitive behavior of social systems. Theor Decis 2(2):109–140
23. Senge PM (1990) The fifth discipline: the art and practice of the learning organization. Doubleday, New York
24. Lad BK, Kulkarni MS (2011) Optimal maintenance schedule decisions for machine tools considering the user's cost structure. Int J Prod Res 50(20):5859–5871
25. Levitt J (2011) Complete guide to preventive and predictive maintenance, 2nd edn. Industrial Press Inc., New York
26. Komonen K (2002) A cost model of industrial maintenance for profitability analysis and benchmarking. Int J Prod Econ 79(1):15–31
27. Jokinen T, Ylén P, Pyötsiä J (2011) Dynamic model for estimating the added value of maintenance services. In: Proceedings of international conference of the system dynamics society
28. Korpi E, Ala-Risku T (2008) Life cycle costing: a review of published case studies. Manag Audit J 23(3):240–261

29. Barringer PH (2003) A life cycle cost summary. In: Proceedings of the international conference of maintenance societies
30. Örjan L (2000) TPM: Vägen till ständiga förbättringar. Studentlitteratur, Lund
31. Haarman M, Delahay G (2004) Value driven maintenance—new faith in maintenance. Mainnovation, Dordrecht
32. Woodhouse J (2006) PAS-55—asset management: concepts & practices. In: Proceedings of international maintenance conference
33. Repenning NP, Sterman JD (2001) Nobody ever gets credit for fixing problems that never happened: creating and sustaining process improvement. Calif Manag Rev 43(4):64–88
34. Linnéusson G (2009) On System Dynamics as an Approach for Manufacturing Systems Development. Lic. Technical Report No. 58. Chalmers University of Technology
35. Liker JK, Meier D (2006) The Toyota way fieldbook: a practical guide for implementing Toyota's 4Ps. McGraw-Hill, New York
36. Repenning NP, Sterman JD (2002) Capability traps and self-confirming attribution errors in the dynamics of process improvement. Adm Sci Q 47(2):265–295
37. Jambekar AB (2000) A systems thinking perspective of maintenance, operations, and process quality. J Qual Maint Eng 6(2):123–132
38. Thun J-H (2006) Maintaining preventive maintenance and maintenance prevention: analysing the dynamic implications of Total Productive Maintenance. Syst Dyn Rev 22(2):163–179
39. Gopalakrishnan M, Skoogh A, Laroque C (2014) Simulation-based planning of maintenance activities by a shifting priority method. In: Proceedings of winter simulation conference (WSC)

Investigation of Causes of Mining Machines Maintenance Problems

Ljubisa Papic, Srdja Kovacevic, Diego Galar and Adithya Thaduri

Abstract Human errors in the area of mining engineering are of critical issue that has serious concerns in safety, operation and production performance. There is a need for finding cause and effect relations with respect to the maintenance issues in order to detect, scrutinize and take necessary actions to reduce it. This paper deals with the human errors in the mining machines for the maintenance problems using fishbone cause and effect analysis. The investigation of these causes and effects are carried out during different operating conditions in typical mining industry and potential problems are assessed. There are several recommendations are provided to reduce the effect of human error so as to increase production by careful consideration of maintenance activities.

Keywords Root causes · Maintenance · Quality management

1 Introduction

American scientific influence on quality management improvement was brought by a formation of Japan scientific school in management after the II World War [1]. Typical representatives of this school that have to be mentioned are, before all,

L. Papic (✉)
University of Kragujevac, Kragujevac, Serbia
e-mail: dqmcenter@mts.rs

S. Kovacevic
JP PK Kosovo Obilic, Belgrade, Serbia
e-mail: srdja.kovacevic@mts.rs

D. Galar · A. Thaduri
Luleå University of Technology, Luleå, Sweden
e-mail: diego.galar@ltu.se

A. Thaduri
e-mail: adithya.thaduri@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_21

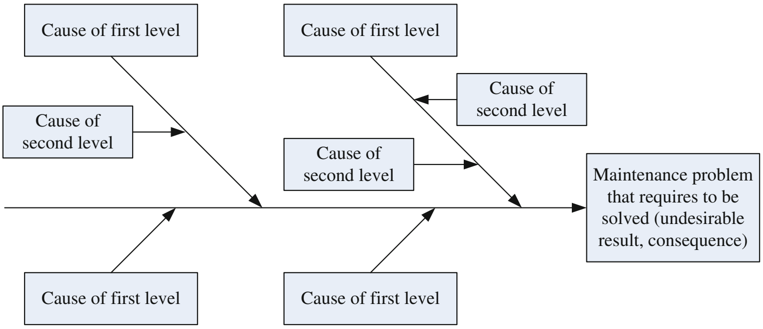


Fig. 1 General schematic for maintenance problem analysis that requires resolving, adopted for the cause investigation of mining machines maintenance problems

Kaoru Ishikawa and Genichi Taguchy who had great impact to the development of statistic methods in quality management [2]. Kaoro Ishikawa is the first, in the world practice, who proposed an original graphic method of cause-effect relations analysis, and it is entitled “Ishikawa diagram” or “cause-effect diagram” or “fish-bone diagram” [3]. It is difficult to find a working area today that doesn`t use the diagram of Kaoru Ishikawa [4] for resolving a quality problem that requires to be resolved, including the maintenance problem. Method of cause-effect diagram represents the method of analysis with the result that establishes which effects are caused by certain causes. General schematic of this method, which will be used in the subject investigation of a problem of mining machines maintenance, is shown on Fig. 1.

The method of cause-effect diagram will be used in the subject investigation to detect and systematize factors (causes) that affect the results performing for the mining machines maintenance operation, i.e. sources that cause a maintenance problem [5]. The task of this qualitative analysis is to undertake corrective or preventive measures to eliminate the problems of mining machines maintenance, after their detection. In that way, the use of cause-effects diagram method would be shown as an effective tool to perform the corrective and preventive measures as mandatory procedure of the integrated management system in the organization for example in Taiwan construction industry [6].

2 Design of Cause-Effects Diagram in the Integrated Management System Analysis of Overhaul Organization

A cause-effects diagram can be formed by either internal or external auditors who perform the verification of an organization in order to establish critical places of the integrated management system in future audits [7]. The order of outlining the cause-effects diagram in the analysis of an integrated management system of the overhaul organization is given in steps 1–5.

Step 1: Quality of the product and work processes, as the main characteristics of the outlet values of an overhaul organization is conditioned by the series of influences of various characters, size, direction and course [8]. These influences are considered by:

- possibilities of managing the quality of products and work processes—represents the conditions of the subject process, and
- deviation of quality of products and work processes from the values anticipated by a design—represents the causes of a potential problem.

For further procedures it is possible to formulate problem in the following form: Considered quality indicator is: “Maintenance Problem”, i.e. “Holdup in maintenance”.

Step 2: The main factors—causes related with the selected indicator are defined in this step:

1. Maintenance object.
2. Store house.
3. Power facility.
4. Workers.
5. Maintenance technology.

Step 3: Further on, the factors of the second level are defined:

1.1 Life duration
1.2 Failure modes
………………………
………………………
2.1 Spare parts
2.2 Process scope of documents
………………………
………………………
3.1 Operative conditions of equipment
3.2 Type of equipment
………………………
………………………
4.1 Qualification
4.2 Health
………………………
………………………
5.1 Diagnostics of failures
5.2 Documentation
………………………
………………………

Step 4: Factors of third level are separated for each factor of second level:

1.1.1 Operational condition
………………………
………………………
2.1.1 Delivery of spare parts
………………………
………………………
3.1.1 Life duration
………………………
………………………
4.1.1 Level of training
………………………
………………………
5.1.1 Program of damage detection

Step 5: Cause-effects diagram is formed in this step, on the basis of previously established factors, where:

- structure made of five branches which correspond to the main factors (causes) is selected for the main diagram structure,
- detail breaking of diagram is performed by drawing the factor (cause) lines towards each corresponding branch, and
- procedure of spreading (branching) of the diagram is performed in cases when it is estimated that certain factors (causes) are in cause-effect connection, in a series or parallel way.

The obtained result is shown on Fig. 2.

For the need of an internal audit, certification or supervision audit, this cause-effects diagram can be used in the following way:

1. Formulation of five directions of the verification:
   - maintenance object,
   - storehouse,
   - power equipment,
   - workers,
   - overhaul technology.
2. Organization of an audit for each subsequent direction, in accordance with the classification of the third level factors.

It is important to stress, due to specificity of classification (structuring), that auditors must not interrupt each other, since their audits are based on different starting data, normative documentation, etc. [9].

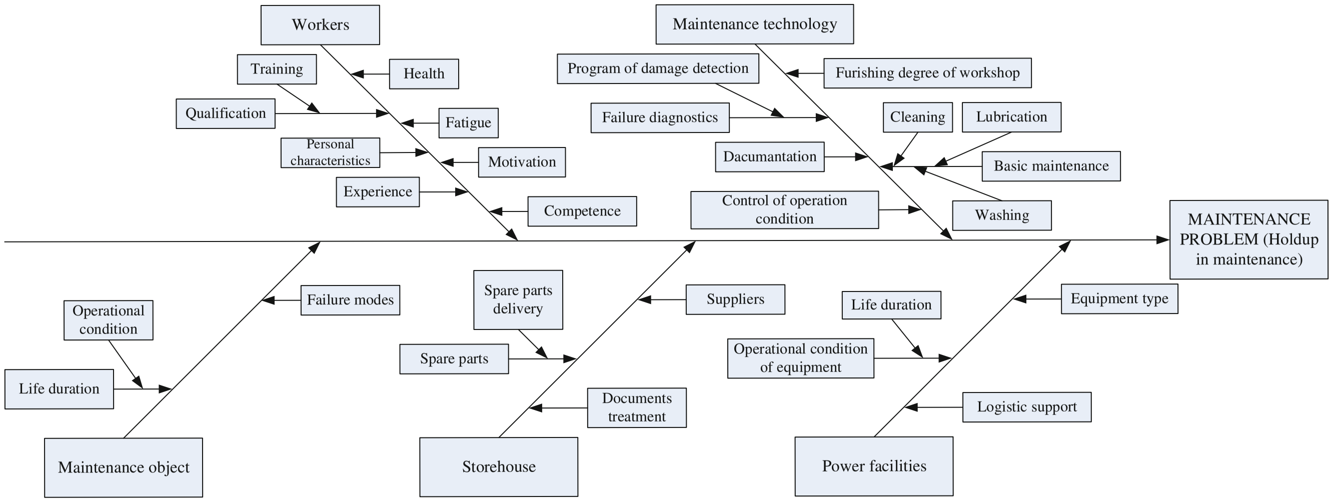


Fig. 2 Cause-effects diagram in analysis of integrated management system of overhaul organization

In case that internal audits, certification and supervision audits include the verification of reliability indicators, it would be useful to shape the cause-effects diagram that examines “Reliability problem (mean UP TIME)” indicator [10].

3 Investigation of Causes of Personnel Errors During Performing of Mining Machines Maintenance Operations

Investigation of causes of human errors during performing the mining machines maintenance operations is performed by a team work in the regime of Brainstorming method. The team would have to act in accordance with all the recommendations for the organization of Brainstorming [11]. The main recommendations are for: team composition, working way in the team, role of the team leader. The team generated ideas about causes of the maintenance problem that requires being resolved.

The rule, appropriate for making a starting (general) cause-effects diagram, which is applicable in most of real situations, is applied in the subject investigation. This rule anticipates that there always exists certain number of categories of possible causes to some consequences (undesirable results) of work process.

In resolving a particular maintenance problem, the investigation revealed the factors (causes) on which the undesirable result or consequence depends [12–15]:

> “Human error with the highest degree of risk during performing of mining machine maintenance operations”.

The investigation firstly determined and separated five causes, in the sense as shown on Fig. 3:

- lack of training,
- inappropriate information,
- lack of experience,

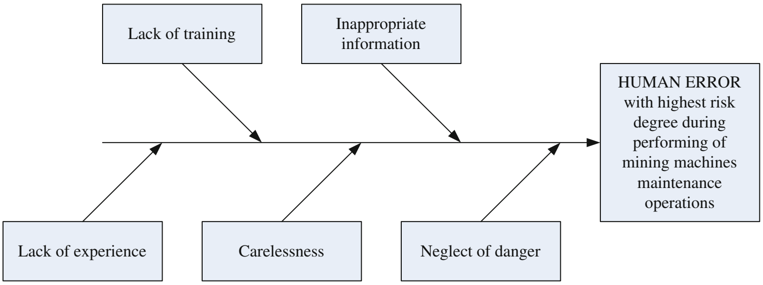
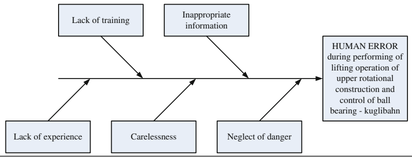
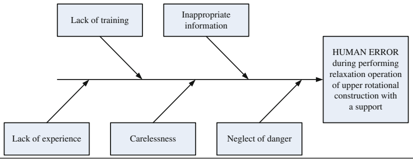
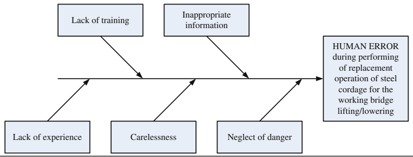
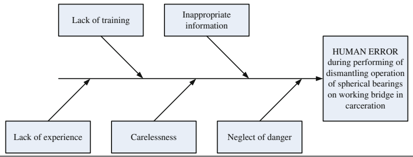
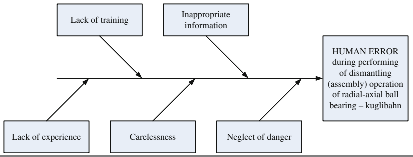
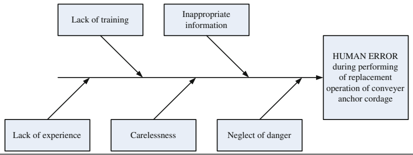
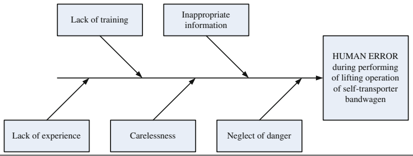
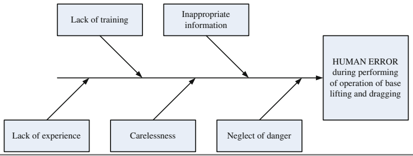

Lack of training
Inappropriate information
HUMAN ERROR with highest risk degree during performing of mining machines maintenance operations
Lack of experience
Carelessness
Neglect of danger

Fig. 3 Potential causes of human errors with the highest risk degree during performing of mining machines maintenance operations in the form of cause-effects diagram

Table 1 Recapitulation of types and causes of human errors that have the highest risk degree during performing of corresponding maintenance operations of various types of mining machines

Type of mining machines	Type of maintenance operation	Types and causes of human errors with the highest risk degree at performing of mining machines maintenance operations
Bucket wheel excavator	Lifting operation of the upper rotational construction and control of ball bearing—kuglbahn	 Lack of training Inappropriate information HUMAN ERROR during performing of lifting operation of upper rotational construction and control of ball bearing - kuglibahn Lack of experience Carelessness Neglect of danger
	Relaxation operation of upper rotational construction with a support	 Lack of training Inappropriate information HUMAN ERROR during performing relaxation operation of upper rotational construction with a support Lack of experience Carelessness Neglect of danger
	Replacement operation of steel cordage for the working bridge lifting/lowering	 Lack of training Inappropriate information HUMAN ERROR during performing of replacement operation of steel cordage for the working bridge lifting/lowering Lack of experience Carelessness Neglect of danger
	Dismantling operation of spherical bearings on working bridge incarceration	 Lack of training Inappropriate information HUMAN ERROR during performing of dismantling operation of spherical bearings on working bridge in carceration Lack of experience Carelessness Neglect of danger
Landfill machine	Dismantling (assembly) operation of radial-axial ball bearing—kuglbahn	 Lack of training Inappropriate information HUMAN ERROR during performing of dismantling (assembly) operation of radial-axial ball bearing – kuglibahn Lack of experience Carelessness Neglect of danger
Damping machine	Replacement operation of conveyer anchor cordage	 Lack of training Inappropriate information HUMAN ERROR during performing of replacement operation of conveyer anchor cordage Lack of experience Carelessness Neglect of danger
Self-transporter bandwagon	Lifting operation of self-transporter bandwagon	 Lack of training Inappropriate information HUMAN ERROR during performing of lifting operation of self-transporter bandwagen Lack of experience Carelessness Neglect of danger
Dragline dredge	Operation of base lifting and dragging	 Lack of training Inappropriate information HUMAN ERROR during performing of operation of base lifting and dragging Lack of experience Carelessness Neglect of danger

- carelessness,
- danger neglect.

Investigations performed in the course of the subject work, with recapitulation shown in Table 1, have shown which types and causes of human errors have the highest risk degree at execution of corresponding maintenance operations of various mining machines, as:

- bucket wheel excavator,
- landfill machine,
- dumping machine,
- self-transporter bandwagon,
- dragline dredge

The obtained results, in the sense of potential causes of first level human errors, at execution of the following maintenance operations:

- lifting operation of upper rotational construction and control of ball bearing-kuglbahn,
- relaxation operation of upper rotational construction with a support,
- replacement operation of steel cordage for the working bridge lifting/lowering,
- dismantling operation of spherical bearings on working bridge incarceration,
- dismantling (assembly) operation of radial-axial ball bearing—kuglbahn,
- replacement operation of conveyer anchor cordage,
- lifting operation of self-transporter bandwagon,
- operation of base lifting and dragging, are shown on Figs. 4, 5, 6, 7, 8, 9, 10 and 11.

Since several (five) main factors (causes of first level human errors) are revealed in the subject investigation, cause-effects diagrams shown on Figs. 4, 5, 6, 7, 8, 9, 10 and 11 are furthermore treated in detail for certain first level causes, performing the investigation of causes of human errors of second and higher levels.

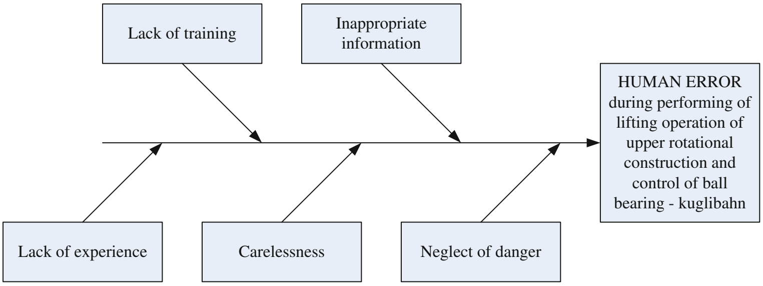

Lack of training
Inappropriate information
HUMAN ERROR during performing of lifting operation of upper rotational construction and control of ball bearing - kuglibahn
Lack of experience
Carelessness
Neglect of danger

Fig. 4 Potential causes of human error during performing of lifting operation of upper rotational construction and control of ball bearing—kuglbahn

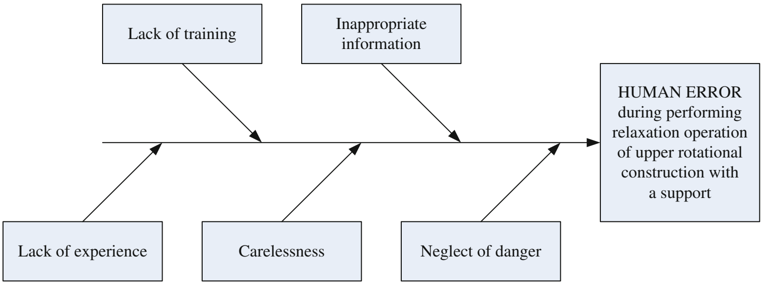

Lack of training
Inappropriate information
HUMAN ERROR during performing relaxation operation of upper rotational construction with a support
Lack of experience
Carelessness
Neglect of danger

Fig. 5 Potential causes of human error during performing of relaxation operation of upper rotational construction with a support

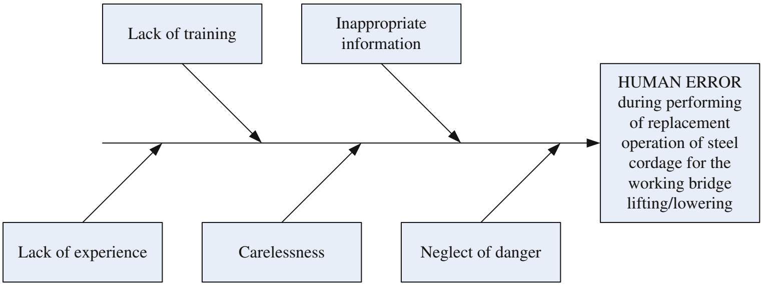

Lack of training
Inappropriate information
HUMAN ERROR during performing of replacement operation of steel cordage for the working bridge lifting/lowering
Lack of experience
Carelessness
Neglect of danger

Fig. 6 Potential causes of human error during performing of replacement operation of steel cordage for the working bridge lifting/lowering

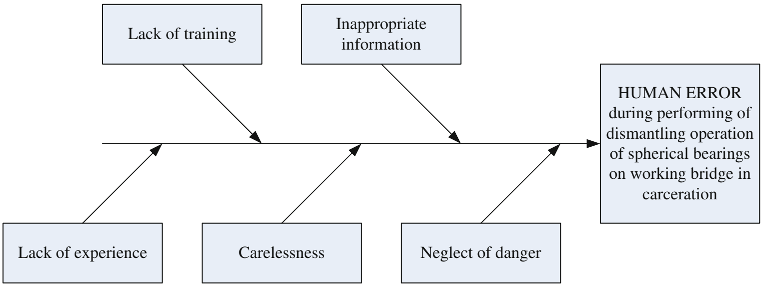

Lack of training
Inappropriate information
HUMAN ERROR during performing of dismantling operation of spherical bearings on working bridge in carceration
Lack of experience
Carelessness
Neglect of danger

Fig. 7 Potential causes of human error during performing of dismantling operation of spherical bearings on working bridge in carceration

Further solution of the subject problem in a qualitative way established the causes of the second and higher levels that generate the first level causes of human errors with the highest risk degree, during performing of mining machines maintenance operations. The causes are connected to the following:

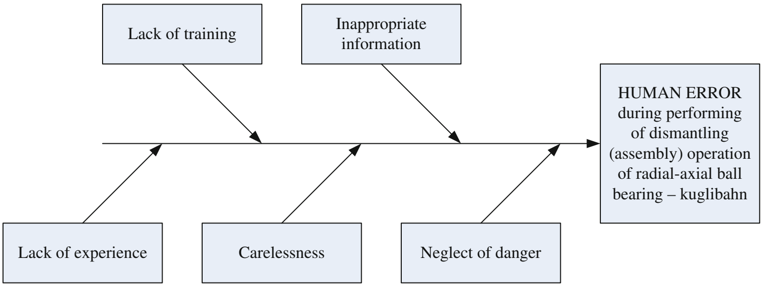

Lack of training
Inappropriate information
HUMAN ERROR during performing of dismantling (assembly) operation of radial-axial ball bearing – kuglibahn
Lack of experience
Carelessness
Neglect of danger

Fig. 8 Potential causes of human error during performing of dismantling (assembly) operation of radial-axial ball bearing—kuglbahn

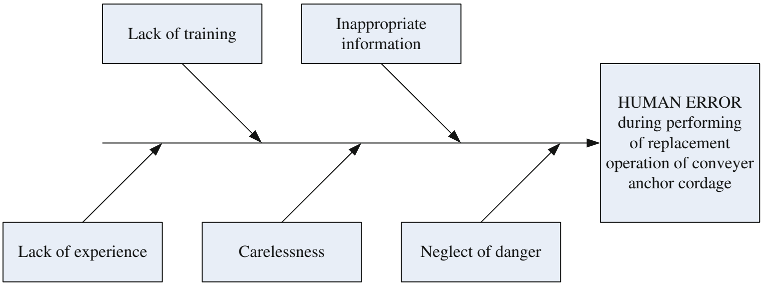

Lack of training
Inappropriate information
HUMAN ERROR during performing of replacement operation of conveyer anchor cordage
Lack of experience
Carelessness
Neglect of danger

Fig. 9 Potential causes of human error during performing of replacement operation of conveyer anchor cordage

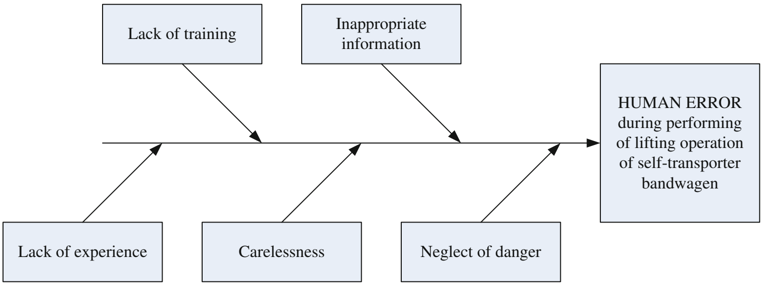

Lack of training
Inappropriate information
HUMAN ERROR during performing of lifting operation of self-transporter bandwagen
Lack of experience
Carelessness
Neglect of danger

Fig. 10 Potential causes of human error during performing of lifting operation of self-transporter

- lack of training (first level cause), Fig. 12,
- inappropriate information (first level cause), Fig. 13,
- lack of experience (first level cause), Fig. 14,
- carelessness (first level cause), Fig. 15,

neglect of danger (first level cause), Fig. 16.

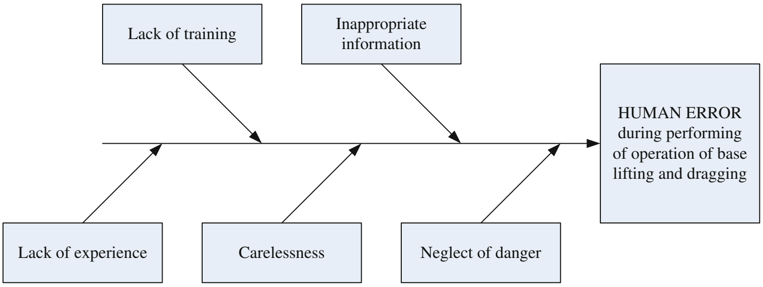

Lack of training
Inappropriate information
HUMAN ERROR during performing of operation of base lifting and dragging
Lack of experience
Carelessness
Neglect of danger

Fig. 11 Potential causes of human error during performing of operation of base lifting and dragging

4 Comment in Research Results

Cause-effect analysis of human errors with the highest risk degree, during performing of mining machine maintenance operations, enabled the division of the main causes (cause of the first level) on less significant causes. Also, this analysis enabled the visual (graphic) presentation of revealed causes and overview of their inter-connection. The results obtained in the performed investigation enable forming of the following partial conclusions:

- Each problem of human factor in mining machines maintenance was investigated from the widest point of view, taking into consideration inner as well as external factors.
- Each proposal about the influential factors or causes of the investigated human factor problems in mining machines maintenance was introduced in a particular location on the paper on which the cause-effect diagram was presented. That location was proposed by the author of the idea. The decision was made by majority of the team members during the Brainstorm process.
- The team members have discussed each cause-effects diagram after its completion. The consultations were also made with the specialists who were not members of the team, but were employees of the open mine pits “Kosovo”—Obilic Company. They provided questions and propositions which were discussed afterwards.
- The copies of cause-effects diagram were shared to the employees of open mine pits “Kosovo”—Obilic Company in order to discuss the diagrams and to obtain their propositions, since they are dedicated to resolve significant problems during the mining machines maintenance in the company.
- The team that performed the investigation involved the direct executors at the workplaces: operators, maintainers, controllers, etc. They are familiar with the maintenance problems from the “inside” and they proposed effective measures for their solution.

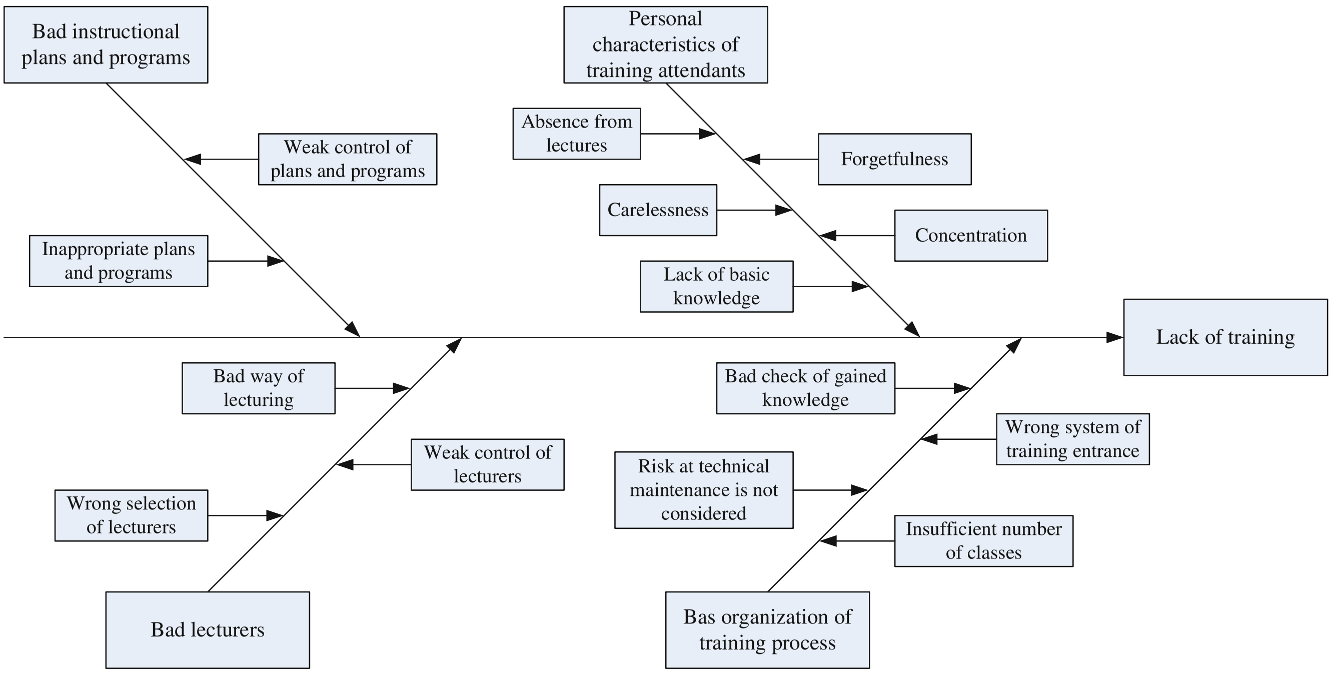

Bad instructional plans and programs
Weak control of plans and programs
Inappropriate plans and programs
Personal characteristics of training attendants
Absence from lectures
Forgetfulness
Carelessness
Concentration
Lack of basic knowledge
Lack of training
Bad way of lecturing
Weak control of lecturers
Wrong selection of lecturers
Bad lecturers
Bad check of gained knowledge
Wrong system of training entrance
Risk at technical maintenance is not considered
Insufficient number of classes
Bas organization of training process

Fig. 12 Cause-effects diagram at analysis of human error causes of first level: “Lack of training”

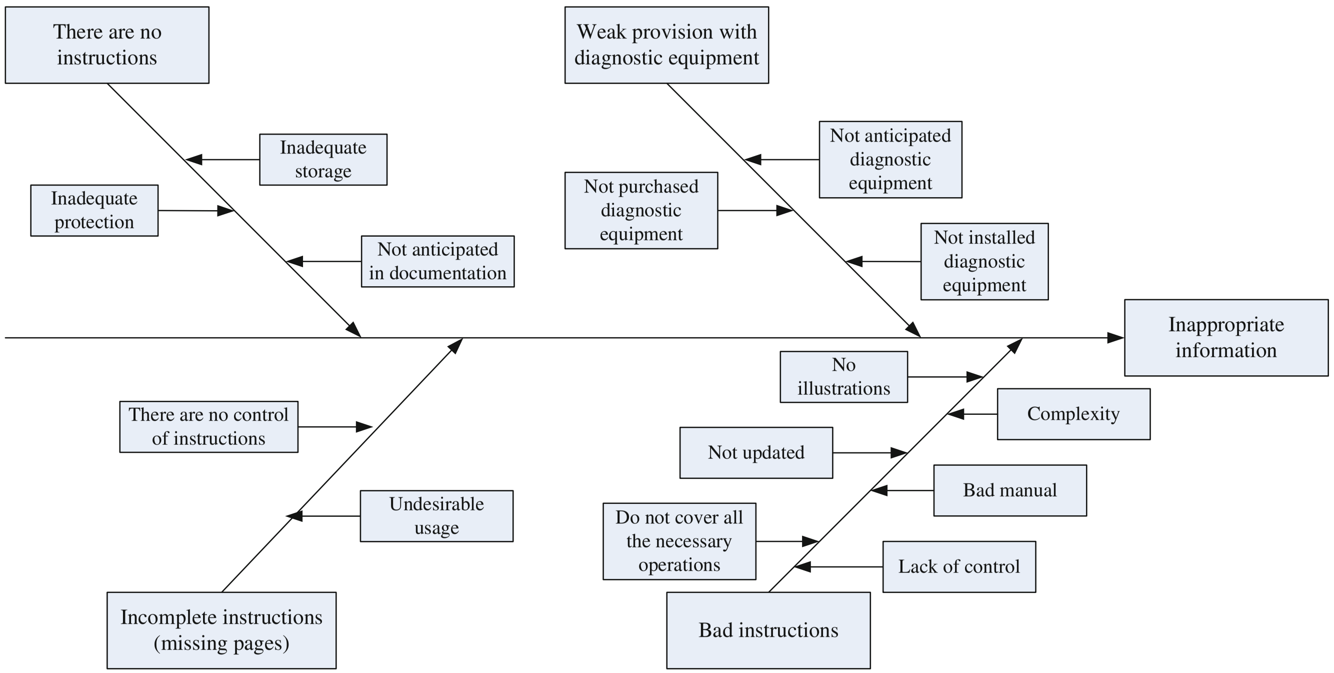

There are no instructions
Inadequate storage
Inadequate protection
Not anticipated in documentation
Weak provision with diagnostic equipment
Not anticipated diagnostic equipment
Not purchased diagnostic equipment
Not installed diagnostic equipment
Inappropriate information
There are no control of instructions
Undesirable usage
Incomplete instructions (missing pages)
No illustrations
Complexity
Not updated
Bad manual
Do not cover all the necessary operations
Lack of control
Bad instructions

Fig. 13 Cause-effects diagram at analysis of human error causes of the first level: “Inappropriate information”

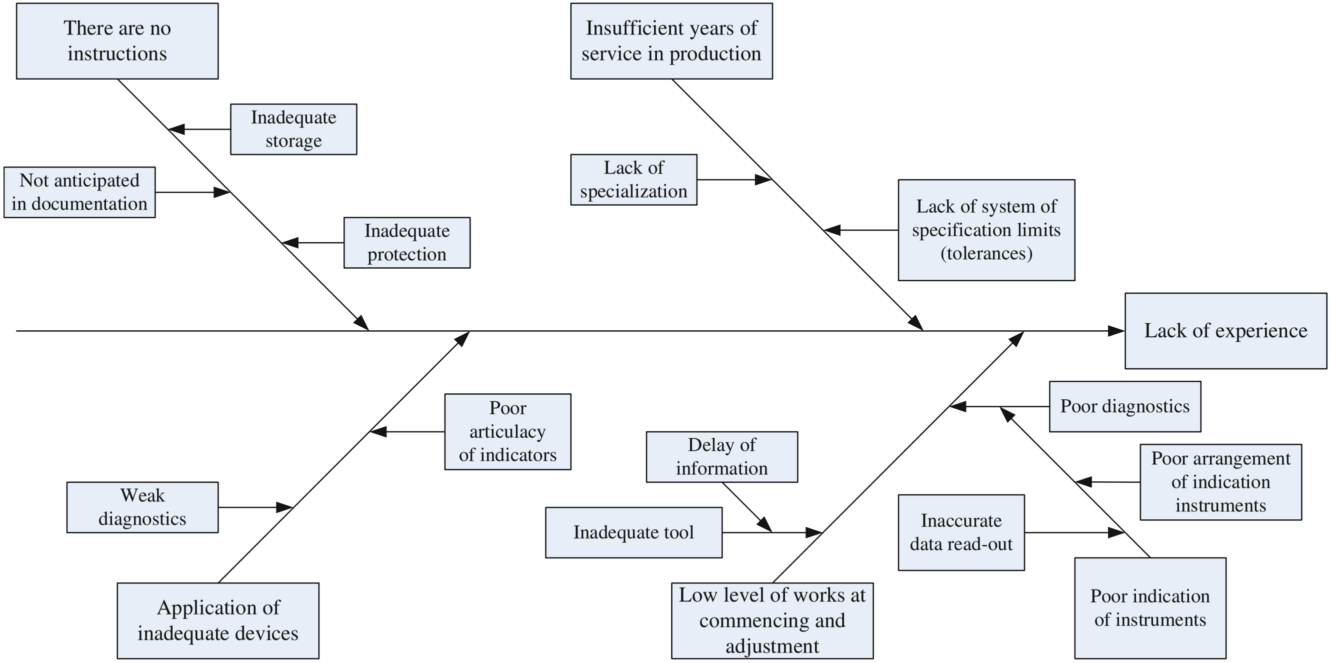

There are no instructions
Inadequate storage
Not anticipated in documentation
Inadequate protection
Insufficient years of service in production
Lack of specialization
Lack of system of specification limits (tolerances)
Lack of experience
Poor articulacy of indicators
Weak diagnostics
Application of inadequate devices
Delay of information
Inadequate tool
Low level of works at commencing and adjustment
Poor diagnostics
Poor arrangement of indication instruments
Inaccurate data read-out
Poor indication of instruments

Fig. 14 Cause-effects diagram at analysis of human error causes of the first level: “Lack of experience”

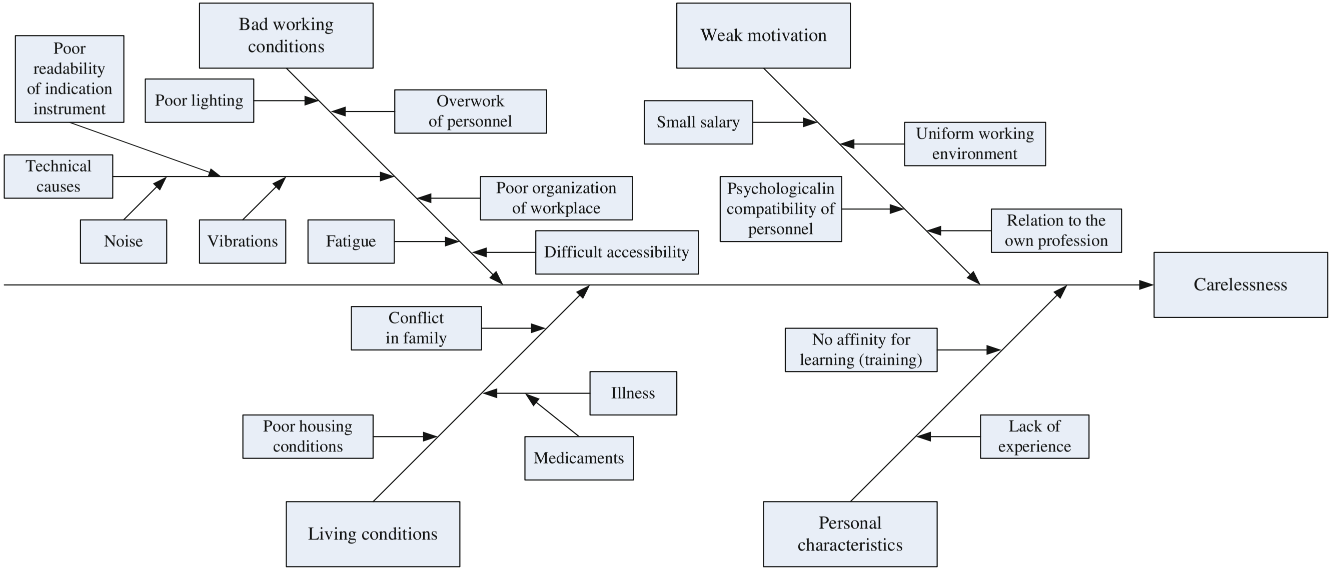


Fig. 15 Cause-effects diagram at analysis of human error causes of the first level: “Carelessness”

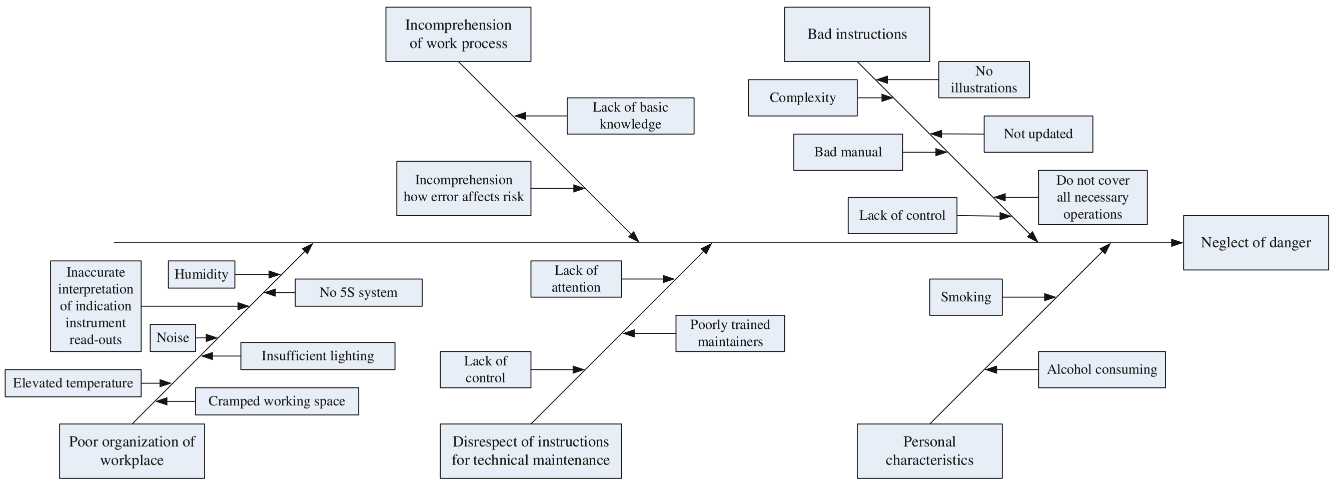


Fig. 16 Cause-effects diagram at analysis of human error causes of the first level: "Neglect of danger"

References

1. Tsutsui WM (2001) Manufacturing ideology: scientific management in twentieth-century. Princeton University Press, Japan
2. Kackar RN (1989) Taguchi's quality philosophy: analysis and commentary. In Quality control, robust design, and the Taguchi method. Springer, US, pp 3–21
3. Ishikawa K, Ishikawa K (1982) Guide to quality control, vol 2. Asian Productivity Organization, Tokyo
4. Ishikawa, K (1985) What is total quality control? The Japanese way, Vol 215. Prentice-Hall, Englewood Cliffs, p 247
5. Herzig K, Zeller A (2011) Mining cause-effect-chains from version histories. In: 2011 IEEE 22nd international symposium on software reliability engineering (ISSRE). IEEE, pp 60–69
6. Cheng CW, Lin CC, Leu SS (2010) Use of association rules to explore cause–effect relationships in occupational accidents in the Taiwan construction industry. Saf Sci 48(4):436–444
7. Arter DR (2003) Quality audits for improved performance. ASQ Quality Press, Milwaukee, p 152
8. Cooper RG (1996) Overhauling the new product process. Ind Mark Manage 25(6):465–482
9. Whalen RC (2008) The subprime crisis—cause, effect and consequences. J Affordable Hous Community Dev Law 17:219–235
10. Balagurusamy E (1984) Reliability engineering. Tata McGraw-Hill Education, New Delhi
11. Tague NR (2005). The quality toolbox, vol 600. ASQ Quality Press, Milwaukee, p 583
12. Paul PS, Maiti J (2007) The role of behavioral factors on safety management in underground mines. Saf Sci 45(4):449–471
13. Horberry T, Burgess-Limerick R, Steiner LJ (2010) Human factors for the design, operation, and maintenance of mining equipment. CRC Press, USA
14. Dhillon BS, Liu Y (2006) Human error in maintenance: a review. J Qual Maintenance Eng 12(1):21–36
15. Galar D, Stenström C, Parida A, Kumar R, Berges L (2011) Human factor in maintenance performance measurement. In: IEEE international conference on industrial engineering and engineering management (IEEM). IEEE, Piscataway, pp 1569–1576

Part IV
Maintenance Modeling and Analysis

Safety and Availability Evaluation of Railway Signalling Systems

Amparo Morant, Anna Gustafson and Peter Söderholm

Abstract The purpose of this paper is to evaluate the safety and availability of railway signalling systems using Markov models. Since a failure of the signalling systems still allows operation of the railway, it is not sufficient to study their safety and availability by considering only the failures and delays. The safety and availability are evaluated, handling both repairs and replacements by using a Markov model. The model is validated with a case study of Swedish railway signalling systems with different scenarios. The results obtained show that the probability of being in a state where operation is possible in a degraded mode is greater than the probability of not being operative at all, which reduces delays but requires other risk mitigation measures to ensure safe operation.

Keywords Railway · Signalling systems · Operation · Maintenance · Availability · Reliability · Safety · Markov · Dependability · RAMS

A. Morant (✉)
Luleå Railway Research Center, Operation and Maintenance Engineering,
Luleå University of Technology, Luleå, Sweden
e-mail: Amparo.morant@ltu.se

A. Gustafson
Mining and Geotechnical Engineering, Luleå University of Technology,
Luleå, Sweden
e-mail: Anna.gustafson@ltu.se

P. Söderholm
Trafikverket, Quality Technology and Management, Luleå University of Technology,
Luleå, Sweden
e-mail: Peter.soderholm@trafikverket.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_22

1 Introduction

The railway can be divided into different systems, such as the rolling stock, the track, the power supply, the signalling system, etc. based on their functionality [1]. Signalling systems play an important role in the control, supervision and protection of rail traffic. Their functionality is based on the principle of "fail safe", which means that the railway section where a failure is located will not be fully operative, until the failure is repaired, to ensure safety.

The operation of a signalling system is based on the interoperability of its different systems. Hence, the availability of these systems directly affects the capacity of the whole railway network. The need to assure interoperability between different parts to obtain the desired output defines a system of systems (SoS) [2, 3]. Railway signalling systems are considered to be a SoS. The safety and availability are evaluated during the maintenance and operation phases of the life cycle, handling both repairs and replacements of the different systems of the various assets comprising the SoS. When managing SoS, it is not possible to consider the different parts independently; functionality depends on the relationship between them [4] Furthermore, the complex architecture of electronics and the interdependency between the components and systems make it difficult to identify and analyse anomalous behaviours [5]. In the case of signalling systems, this difficulty may be illustrated by the number of no fault found failures or not defined failures that are recorded; these represent up to 70 % of the total number of work orders [6]. Since a failure of the signalling systems still allows operation of the railway, albeit limited, it is not sufficient to study their effect on the railway operation in terms of reliability and safety by considering only the failures and delays.

A failure in a signalling system has economic consequences (penalties, high amount of maintenance resources, etc.), can affect the operation (delays, cancellations, speed restrictions, etc.) and have safety consequences. With a failed signalling system, the train will operate in a degraded mode, with safety assured by other mitigation measures, such as low speed restrictions. The possibility of operating in a degraded mode reduces the economic and operational effects of a failure of the signalling systems, but makes it more difficult to evaluate the railway operation, since a failure will not necessarily be visible when considering the delays or cancelations, even though safety has been compromised.

Some research has been done on the area of the reliability, availability, maintainability and safety (RAMS) of railway signalling systems during its operation and maintenance life cycle phase. Tao [7] presented a two-stage safety analysis model for railway level crossing surveillance systems by using Fuzzy Petri Nets, Fault Tree Analysis and a Markov model to handle the incomplete safety-related data; Tao also provided an empirical study assessing the safety status of a level crossing surveillance system. Anik et al. [8] compared the different system architectures at the algorithm level in terms of the safety integrity level of a railway station by using failure trees and Markov processes. Tan et al. [9] developed a Markov model to evaluate the RAMS of the system depending on the architecture redundancy of the trains' vital computer. Brkic and Adamovic [10] presented a Markov model that can express the reliability of a signalling system and evaluate the significance of the system's individual elements both qualitatively and quantitatively; the model was validated using a combination of real data from maintenance records and estimations where the data were not sufficiently precise. Kohlik and Kubatova [11] stated that dependability models allow calculating the rate of an event leading to a hazard state, which can result in material loss, serious injuries or casualties; they used a hierarchical dependability model based on Markov chains to speed up the hazard rate calculation. Bondavalli et al. [12] developed a model based on discrete time Markov chains combined with stochastic activity networks methodologies and applied hierarchical modelling to perform a dependability analysis of the safety nucleus subsystem of a railway interlocking. The previous contributions all focused on the evaluation of a particular system. In contrast, this paper evaluates the whole SoS of signalling systems. This is a good approach to use when it is not possible to determine the failed system or failure mode. This paper is also based on records from corrective maintenance showing the variance found in real data; these records allow us to confirm the validity of the model for future implementation in industry.

Various authors have evaluated the availability and/or safety of railway signalling systems: Markov Chains [13], Monte Carlo Simulation [14] and Stochastic Petri Nets [15, 16] are suitable approaches for stochastic modelling to evaluate the RAMS of a railway signalling system. The Markov model allows handling both repairable and non-repairable systems, hence is appropriate to evaluate the performance of railway signalling systems.

The purpose of this paper is to evaluate the safety and availability of railway signalling systems using Markov models. The knowledge gained will facilitate the decision-making process when improving or updating the railway infrastructure.

2 Research Methodology

The model proposed in this paper is based on the fusion of different types of information obtained from corrective maintenance data records, operational data, and railway architecture. The model can be used when studying the effect of a failure in the SoS of signalling systems on the overall railway operation in terms of safety and availability. The research is based on data obtained from the Swedish Infrastructure Manager (Trafikverket) for a fully operative railway corridor where the ATC (Automatic Train Control) signalling system supervises and controls the network. Corrective maintenance Work Orders (WO) from a determined railway corridor (divided into several track sections) were gathered and processed from the corrective maintenance database (0felia), while the architecture of the railway corridor was obtained from the asset management database (BIS).

No changes of configuration were made during the years included in the maintenance data used for this research on the railway corridor considered. Hence, it can be assumed that the WOs represent maintenance and not design changes or updates. Previous research related to the railway signalling systems provided current theories and suggested ways to improve the dependability of signalling systems, while Trafikverket documentation and unstructured interviews with experts facilitated our understanding of the information and results.

The collected data and information are processed and combined for the analyses, with Excel 2010, Matlab 2014a and the R software (version 3.0.0) used for data processing, model development and validation. The model is based on a Markov process with discrete states and continuous time and is used to measure the probability of the different operational states (safe operation, not operative or operative in degraded modes) of a track section, identifying the systems that most affect a safe operation of the railway. Depending on which system is affected by the failure and the operational status of the railway, the model considers different operational states. Various scenarios are considered to validate the model, including mean values, worst and best case scenario.

2.1 Analysis

The Markov approach is applicable when handling both repairable and non-repairable systems, under the following assumptions [17]:

- The behaviour of the system must be characterised by a lack of memory; that is, the future states of a system are independent of all past states except the immediately preceding one:

$$\mathrm{P}(\mathrm{qn}|\mathrm{qn}-1, \mathrm{qn}-2, \ldots, \mathrm{q}1) = \mathrm{P}(\mathrm{qn}|\mathrm{qn}-1). \tag{1}$$

- The process must be stationary (i.e. the probability of making a transition from one given state to another is the same at all times in the past and future).
- Finally, it must be possible to define the different states of the system.

The transition rates from one state into another can be defined as in Eq. 2, and the transition between the different states of the Markov model is given by the failure, restoration and waiting rates (λ, μ_o and μ_w respectively) of each considered system. The transition rates describe not only the reliability of the process and the design of the components, but also the effectiveness of operation and maintenance practices [18], shown as:

$$\mathrm{Tr.\,rate} = \frac{\text{num. of times a transition occurs from a given state}}{\text{time spent in the given state}} \tag{2}$$

With respect to the transition rate, three time parameters can be defined. The mean operating time between failures (MTBF) is the expectation of the operating time between failures and can be calculated following Eq. 3; the mean time to maintain (MTTM) is the expectation of the time to restore (see Eq. 4), and the mean waiting time (MWT) is the time from the start of the downtime until the driver is allowed by the dispatcher to continue operation in a degraded operating mode:

$$\mathrm{MTBF} = \frac{\text{Total operative time*Nr. of systems}}{\text{num. of failures}} \tag{3}$$

$$\mathrm{MTTM} = \frac{\text{Total Downtime}}{\text{num of failures}} \tag{4}$$

From Eq. 2, the transition between the different states of the Markov model is given by λ, μ_o and μ_w of each system considered (see Eqs. 5, 6 and 7). In particular, μ_w measures the rate of systems staying in the non-operative state.

$$\lambda = \frac{1}{\mathrm{MTBF}} \tag{5}$$

$$\mu_o = \frac{1}{\mathrm{MTTM}} \tag{6}$$

$$\mu_w = \frac{1}{\mathrm{MWT}} \tag{7}$$

2.2 *Analysed Scenarios*

In order to show the probabilities of being in the different operational states obtained by the mean failure and restoration rates, five scenarios are considered and described in Table 1. Scenarios F-1 to F-5 are based on real data gathered from the maintenance databases, from which the MTBF and MTTM have been obtained for the different track sections that compose the railway corridor of the case study.

Scenario F-1 represents the mean values obtained from the corrective maintenance data for the track sections on the studied railway corridor. Scenarios F-2 and F-3 makes it possible to study the effect of the different RAMS variables (such as the MTBF and the MTTM) on the railway operation to see which one has the most influence on a safe operation. Scenarios F-4 and F-5makes it possible to look at the

Table 1 Scenarios to model

	Description
F-1	Mean values of the MTBF and MTTM
F-2	Mean values of the MTBF and 75 % quartile of the MTTM
F-3	25 % quartile of the MTBF and mean values of the MTTM
F-4	Worst case scenario: 25 % quartile of the MTBF and 75 % quartile of the MTTM
F-5	Best case scenario: 75 % quartile of the MTBF and 25 % quartile of the MTTM

variance between the probability states obtained for the worst case scenario and the best case scenario observed from the recordings for all track sections on the railway corridor, looking at the range of values for the MTBF and MTTM (i.e. the lowest reliability and highest maintainability).

3 Case Study

To maximise the capacity of the railway corridor while ensuring safety, the railway signalling system divides the railway corridor into track sections (or blocks) where only one train is allowed at a given time [8]. Figure 1 shows the Reliability Block Diagram (RBD) for the minimum operative section from the point of view of the SoS of signalling systems on a railway network.

The SoS of the signalling system is composed of the following systems [19]:

- Traffic management system (TMS): creates an interface between the traffic operator and the railway network.
- Interlockings (IXL)/Radio Block Centre (RBC): receive the input from the different systems (e.g. track circuits, level crossings, signals, TMS), and calculate and return as an output the train operation restrictions to ensure safe traffic operation.
- Track circuits (TC): enable localisation of the train.
- Balise group (BG): give input from the track to the onboard signalling system (e.g. speed limits, driving mode, etc.).
- Level crossings (LC): coordinate the road traffic crossing the railroad.
- Signals: give or restrict permission to a train on coming into a track section.
- Signalling boards: give the driver fixed information (e.g. on tunnels, bridges, speed restriction areas, etc.).

To ensure safe operation, a track section is supervised by an interlocking located at the end of that section, usually at a station. Signals are placed at the entrance of every section and sometimes in the middle to allow or restrict the passing of a train into that section. Signals restrict the passing of a train when a failure occurs on a track circuit or an interlocking, and warns it to circulate with caution when there is a failure in a level crossing. When a signal fails, the balise group associated with it will force the train to stop. If a balise does not work properly, it will produce an emergency brake (EB). A single TMS controls the railway traffic of various corridors simultaneously. If the TMS fails, the operation has an automatic mode that allows normal operation for a maximum of 2 h. After that time, operation is not

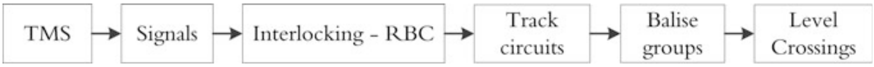


Fig. 1 RBD of a signalling system

possible. If there is a stoppage of operation caused by a failure on the signalling system of a track section, railway operation can still be possible on that section if the dispatcher allows the driver to circulate with caution in a degraded operational mode. In this case, the maximum speed is 40 km/h and the driver's visual supervision is required to ensure safe circulation (e.g. there is no damage in the track; the switch is in the correct position etc.).

The operating time between failures is represented by the total duration of operating time between two consecutive restorations. Signalling systems supervise the railway at all times, not only when a train passes, making them continuously operating items. Therefore, all maintenance time will affect the operation of the signalling system.

3.1 Corrective Maintenance Data

The corrective maintenance data cover WOs from January 2003 until November 2012 on a 203 km long corridor, divided into 50 track sections and located in the northern part of Sweden. Each track section has a different architecture composition for signalling systems. Figure 2 shows the number of systems per track section for the case study. Specifically, 9030 WOs were registered during that period, of which 2455 were associated with signalling systems. The data were processed to eliminate inconsistent or poor-quality records. WOs with a time to restoration equal or less than zero seconds or more than 24 h were discarded (procedures for corrective maintenance establish a WO should be closed after a maximum of 24 h (Trafikverket 2010)). In addition, WOs were discarded if they did not correspond to any track section specified on the architecture database or were related to systems not specified for that track section on the architecture database. This left 1933 WOs. The corrective maintenance data and architecture data were merged for the processing required before modelling. Note: only the failures affecting the operation are accounted for in this model. Hence, the TMS and the signalling boards are beyond the present scope: the TMS is shared by all track sections (even when the WOs are related to a particular section), while the signalling boards do not affect the operation of the railway.

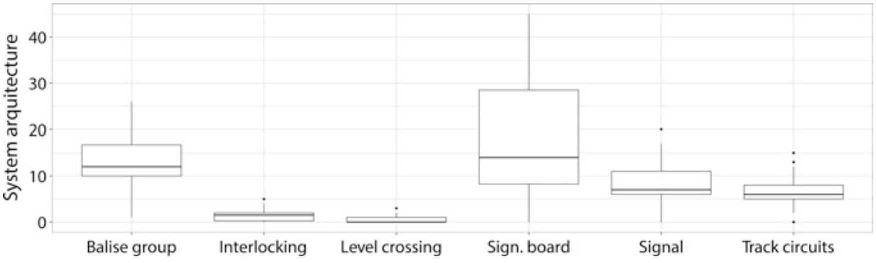


Fig. 2 Number of systems per track section of the studied case (50 track sections)

Table 2 MTBF and MTTM for the case study

	MTBF (Years)			MTTM (h)		
	25 %	Mean	75 %	25 %	Mean	75 %
BG	1.9730	4.7670	9.8631	5.3271	9.2377	12.5167
IXL	0.7182	2.8581	3.0828	4.1839	5.5357	7.2221
LC	0.3846	2.2860	2.4664	2.1171	4.3308	4.7864
Signal	0.8968	2.1464	2.4663	3.1070	5.1494	6.3838
TC	0.8221	2.0040	1.9731	2.0892	3.3577	4.1961

From the case study's corrective maintenance data, it is possible to obtain the information shown in Table 2. Note: mean waiting time is considered to be 5 min (obtained from interviews with experts).

4 Model Development

Depending on the system affected by the failure, the three operational states of the railway infrastructure considered are subdivided, giving a total of 11 states that determine the different operational states and the system affecting operation. The states are described in Table 3. The last two columns of the table show graphically the status of availability and safety, and how these change depending on the state of the railway: green (and happy) face when OK, yellow (and neutral) face when operating in a degraded mode and red (and sad) face when the signalling system is not ensuring safety or the railway is not available.

$$\mathrm{P}=\begin{bmatrix}
1-\lambda_{1,2}-\lambda_{1,4}-\lambda_{1,6}-\lambda_{1,8}-\lambda_{1,10} & \lambda_{1,2} & 0 & \lambda_{1,4} & 0 & \lambda_{1,6} & 0 & \lambda_{1,8} & 0 & \lambda_{1,10} & 0 \\
0 & 1-\mu_{2,3} & \mu_{2,3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mu_{3,1} & 0 & 1-\mu_{3,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1-\mu_{4,5} & \mu_{4,5} & 0 & 0 & 0 & 0 & 0 & 0 \\
\mu_{5,1} & 0 & 0 & 0 & 1-\mu_{5} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1-\mu_{6,7} & \mu_{6,7} & 0 & 0 & 0 & 0 \\
\mu_{7,1} & 0 & 0 & 0 & 0 & 0 & 1-\mu_{7,1} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-\mu_{8,9} & \mu_{8,9} & 0 & 0 \\
\mu_{9,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-\mu_{9,1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-\mu_{10,11} & \mu_{10,11} \\
\mu_{11,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-\mu_{11,1}
\end{bmatrix} \tag{8}$$

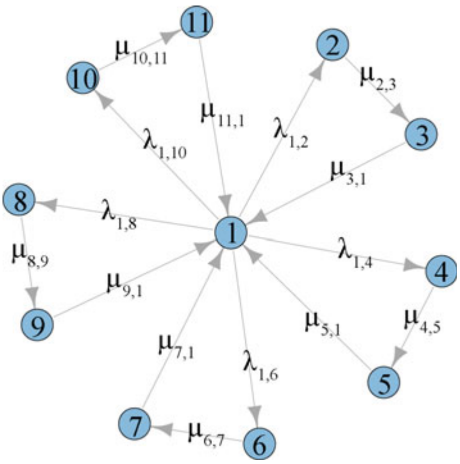
4.1 Model Architecture

The state-space diagram for the Markov process visualised in Fig. 4 shows the different states of the system (see Table 1 for description) and the possible transitions between them. From Fig. 3 it is possible to deduce the simplified stochastic

Table 3 States

States		Description	**Avail.**	**Safety**
St.1	Operative state	The railway operation is possible and the signalling system is fully operative	🙂	🙂
St.2	Faulty—BG failed	The railway operation is not possible and the signalling system is not operative due to a failure on the BG	🙁	🙂
St.3	Degraded—BG failed	The railway operation is possible in degraded mode (40 km/h and the driver is responsible for supervision and protection), the signalling system is not operative due to a failure on the BG	😐	🙁
St.4	Faulty—IXL failed	The railway operation is not possible and the signalling system is not operative due to a failure on the IXL	🙁	🙂
St.5	Degraded—IXL failed	The railway operation is possible in degraded mode (40 km/h and the driver is responsible for supervision and protection), the signalling system is not operative due to a failure on the IXL	😐	🙁
St.6	Faulty—LC failed	The railway operation is not possible and the signalling system is not operative due to a failure on the LC	🙁	🙂
St.7	Degraded—LC failed	The railway operation is possible in degraded mode (40 km/h and the driver is responsible for supervision and protection), the signalling system is not operative due to a failure on the LC	😐	🙁
St.8	Faulty—signal failed	The railway operation is not possible and the signalling system is not operative due to a failure on the signal	🙁	🙂
St.9	Degraded—signal failed	The railway operation is possible in degraded mode (40 km/h and the driver is responsible for supervision and protection), the signalling system is not operative due to a failure on the signal	😐	🙁
St.10	Faulty—TC failed	The railway operation is not possible and the signalling system is not operative due to a failure on the TC	🙁	🙂
St.11	Degraded—TC failed	The railway operation is possible in degraded mode (40 km/h and the driver is responsible for supervision and protection), the signalling system is not operative due to a failure on the TC	😐	🙁

Fig. 3 Markov diagram



transitional probability matrix shown in Eq. 8. The system is considered to be fully operative for the initial state expressed as:

$$\mathrm{P}(\mathrm{t}=0)=(1,0,0,0,0,0,0,0,0,0,0,0)$$

5 Results and Discussion

A Markov model is developed to measure the probability of the three operational states (operative, faulty or degraded) of the railway signalling system, identifying the systems that most affect the safe operation of the railway. The developed model was used to measure the operational effects of dependability improvements of different signalling assets and then validated with a case study of a Swedish railway signalling system using a number of scenarios. Certain assumptions were made, for example, that the failure and repair rate follow an exponential distribution. The Markov model is a tool for maintainers to use when evaluating the safety and availability of the railway operation by analysing the times when operation is possible but the signalling system is not ensuring safety; it also allows simulating the effects of different RAMS improvements. Its simplicity allows the maintainers to use the actual maintenance records to obtain an estimation of the level of safety and availability, despite the lack of detailed data (which would be needed if implementing a more complex model. Further work can be oriented to investigate better models that can give a better estimation of the probabilities, not depending on the assumptions taken in this paper.

Since this research is based on the corrective maintenance affecting the operation (supervision, protection, control, information) recorded in the database 0felia, it does not take into account corrective maintenance that could have been done but has not been recorded, e.g. during inspections. The use of real maintenance data makes the research process more complex, but renders the results more relevant since they reflect the complexity of reality. To increase the quality of the maintenance records (for example, recording the component affected, reducing the "not defined" failure modes or recording if the failure affects the railway operation) would increase the reliability of the obtained results.

Table 4 shows the probabilities of being in the different states for scenarios F-1 to F-5 and uses colours to illustrate the relative probabilities between states, with green being desirable and red undesirable. For the operative state, it is desired to achieve a probability state that is as high as possible, but for faulty and degraded states, the lowest one will give the best results for safety and availability.

Figure 4 shows graphically the state probabilities. This allows to visually comparing the results, looking at the performance variance in a real track section. In the figure, the difference between the values obtained for the state probabilities for the various scenarios are shown. For example, for the LC, the state probability of being in a degraded state obtained for scenario F-4 is 14 times the one obtained for the F-5

Table 4 Probabilities of being at different states for the real data scenarios (F-1 to F-5) (*10^−2)

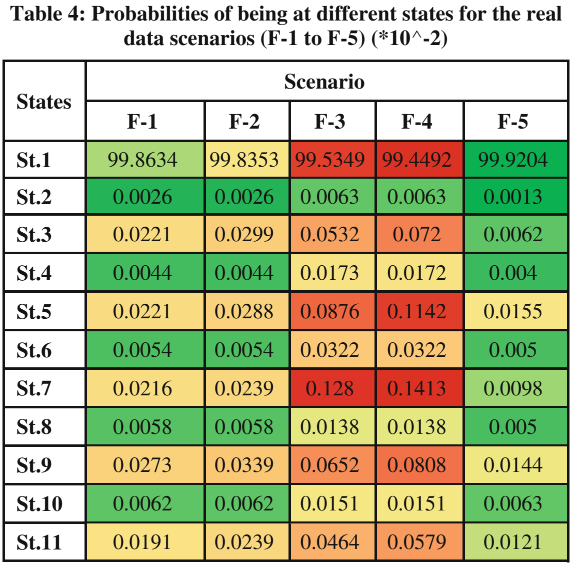


Table 4: Probabilities of being at different states for the real data scenarios (F-1 to F-5) (*10^-2)

States	Scenario				
	F-1	F-2	F-3	F-4	F-5
St.1	99.8634	99.8353	99.5349	99.4492	99.9204
St.2	0.0026	0.0026	0.0063	0.0063	0.0013
St.3	0.0221	0.0299	0.0532	0.072	0.0062
St.4	0.0044	0.0044	0.0173	0.0172	0.004
St.5	0.0221	0.0288	0.0876	0.1142	0.0155
St.6	0.0054	0.0054	0.0322	0.0322	0.005
St.7	0.0216	0.0239	0.128	0.1413	0.0098
St.8	0.0058	0.0058	0.0138	0.0138	0.005
St.9	0.0273	0.0339	0.0652	0.0808	0.0144
St.10	0.0062	0.0062	0.0151	0.0151	0.0063
St.11	0.0191	0.0239	0.0464	0.0579	0.0121

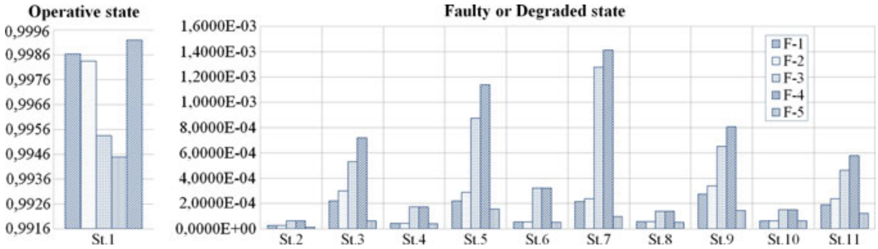


Fig. 4 Probabilities of the railway of being on the different states for the real data scenarios (F-1 to F-5)

scenario. This difference is also remarkable for the BG (11 times). The differences are minor for faulty states, even though they remain identifiable: six times for the faulty state linked to the LC and five times for the one linked to the BG.

Operating the railway in a degraded state can reduce the delays caused by a failure of the signalling system, but this does not change the fact that the signalling system is failed and, hence, safety cannot be ensured. In order to evaluate the safety and availability of the railway, we must not only look into reliability and maintainability (i.e. considering failures and delays) but also consider the probability of operation in a degraded mode. From a safety perspective, the better option is a signalling system with lower reliability but a safer design than one with higher reliability but with a higher probability of operating in a degraded mode.

There is also a difference between the probabilities of the different scenarios of the system most affecting the railway operation. For example, for scenarios F-1 and F-2, the maximum probability of being in a faulty state is linked to a failure of the TC, and the maximum probability of being in a degraded state is linked to a failure of a signal. For scenarios F-3 and F-4, the maximum probabilities for both the faulty state and the degraded state are related to the LC. For scenario F-5, the maximum probability of being in a faulty state is related to the TC and to the IXL for a degraded state.

The smallest difference between the state probabilities occurs for the LC and the TC in scenario F-5, where the probability of being in a degraded state is two times higher than the probability of being in a faulty state. The maximum difference is obtained for the BG in scenario F-4, with a probability of being in a degraded state that is 11 times higher than the probability of not being operative.

The differences in the results obtained for scenarios F-1 to F-5 can be linked to the fact that these results are obtained from operational instead of inherent reliability and maintainability data. Hence, other factors related to the environment, operation, etc. can influence the behaviour of the systems. The logistics related to the waiting time for performing corrective maintenance in a certain location also play an important role in the real repair rates. Maintenance improvements can be oriented, for example, to reduce the waiting time related to logistics if it is necessary to reduce the degraded operational mode.

This paper has used the 50, 25 and 75 % quartiles to show the range of variation that can be obtained when implementing the model, depending on the input data. The choice of these values is more for the purpose of easy visibility than anything else. The results of other simulations using the median, absolute maximums and minimums, and 5 and 10 % quartiles showed no relevant differences.

Even though this research has used the case study of the Swedish signalling system to validate the model, it can be generalised to other types of signalling systems or railway networks, as it can be adapted to fit any existing differences.

6 Conclusions

The purpose of this paper is to evaluate the safety and availability of railway signalling systems using Markov models. The following conclusions can be drawn:

- The Markov model is a tool for maintainers to use when evaluating the safety and availability of the railway operation by analysing the times when the operation is possible but the signalling system is not ensuring the safety.
- The results obtained from the model show that the probability of being in a state where operation is possible in a degraded mode is greater than the probability of not being operative at all, which reduces delays but requires other risk mitigation measures to ensure safe operation.

- The model allows the comparison of corrective maintenance data from different locations, architectures or design solutions, thereby assisting in the decision-making process when improving or updating the railway infrastructure. This last point can be the subject of future research.

Acknowledgments The authors gratefully acknowledge the sponsors of ICRESH-ARMS 2015. The authors gratefully acknowledge the financial and intellectual support of Luleå Railway Research Center (JVTC) and the Swedish Transport Administration (Trafikverket).

References

1. Pěnička M (2007) Formal approach to railway applications. Lecture notes in computer science (including subseries Lecture notes in artificial intelligence and Lecture notes in bioinformatics). Springer, pp 504–520
2. Boardman J, Sauser B (2006) System of systems—The meaning of Of. In: Proceedings IEEE/SMC international conference on system of systems engineering, Los Angeles, CA, pp 118–123
3. Gorod A, Sauser B, Boardman J (2008) System-of-systems engineering management: a review of modern history and a path forward. IEEE Syst J 2(4):484–499
4. Brownsword L, Fisher D, Morris E, Smith J, Kirwan P (2006) System-of-systems navigator: an approach for managing system-of-systems interoperability. Software Engineering Institute (SEI)
5. Dorj E, Chen C, Pecht M (2013) A bayesian hidden markov model-based approach for anomaly detection in electronic systems. In: IEEE aerospace conference proceedings
6. Morant A, Larsson-Kråik P-O, Kumar U (2014) Data-driven model for maintenance decision support—a case study of railway signalling systems. Inst Mech Eng. Proc. Part F: J Rail Rapid Transit. doi:10.1177/0954409714533680
7. Tao C (2009) A two-stage safety analysis model for railway level crossing surveillance systems. In 2009 IEEE international conference on control and automation, p 1497
8. Anik VG, Ustoglu I, Kaymakci OT (2011) The functional safety calculation of a real interlocking system in Turkey. In: 2011 IEEE international conference on mechatronics, ICM 2011—Proceedings; p 71
9. Tan P, He W, Lin J, Zhao H, Chu J (2011) Design and reliability, availability, maintainability, and safety analysis of a high availability quadruple vital computer system. J of Zhejiang Univ Sci A 12(12):926–935
10. Brkic R, Adamovic Z (2011) Research of defects that are related with reliability and safety of railway transport system. Russ J Nondestr Test 47(6):420–429
11. Kohlik M, Kubatova H. (2013). Markov chains hierarchical dependability models: Worst-case computations, LATW 2013—14th IEEE Latin-American Test Workshop
12. Bondavalli A, Nelli M, Simoncini L, Mongardi G (2001) Hierarchical modelling of complex control systems: dependability analysis of a railway interlocking. Comput Sys Sci Eng 16 (4):249–261
13. Chen H, Qian Y (2012) Reliability and safety analysis of cross-redundant Structure based on Markov process In: Proceedings—2012 5th international symposium on computational intelligence and design, ISCID 2012 p 406
14. Hasanzadeh Z, Sandidzadeh MA (2008).The reliability evaluation of interlocking system for improving the operation factors—a case study in Tehran metro. In: Proceedings of the IASTED international conference on modelling and simulation, p 274

15. Adamyan A, He D (2004) System failure analysis through counters of Petri net models. Qual Reliab Eng Int 20(4):317–335
16. Patra AP, Kumar U (2010) Availability analysis of railway track circuits. Proc of the Inst of Mech Eng, Part F: J Rail and Rapid Transit 224(3):169–177
17. Billinton R, Allan RN (1992) Reliability evaluation of engineering systems—concepts and techniques. Plenum Press, New York, pp 260–308
18. Lipsett MG, Gallardo-Bobadilla R (2013) Modeling risk in discrete multi-state repairable systems, asset condition, information systems and decision models. Eng Asset Manag Rev Springer: pp 187–205. doi: 10.1007/978-1-4471-2924-0_10
19. Trafikverket (2010) Manual—use of 0felia for analysis (Handledning—att använda Ofelia för analytiker). Report, Trafikverket, Borlänge, (in Swedish)

Applying Multi-factorial Pareto Analysis in Prioritizing Maintenance Improvement Initiatives

Marcus Bengtsson

Abstract One practical solution in prioritizing which maintenance improvement initiative to undertake is by using Pareto analysis. This has also been advocated for, as one tool (of many), to use within maintenance by many maintenance research practitioners. However, there are some drawbacks and potential sources of errors if not being cautious. The main objective of this research is to develop and test a tool to be utilized for prioritization purposes of maintenance improvement initiatives. The purpose of the paper is to exemplify, from an industrial case, some of the strengths and weaknesses of this approach.

Keywords Prioritization · Pareto analysis · Maintenance improvements

1 Introduction

Since the choice of possible maintenance improvements in a production process is basically infinite whilst maintenance resources are not it is essential that, for instance, improvement initiatives are prioritized. Basically, improvements in maintenance can be performed from two perspectives: increasing effectiveness in the manufacturing equipment (i.e., effect of maintenance activities) and increasing efficiency in the maintenance work being performed. Maintenance effectiveness is connected to indirect maintenance cost (cost for lost production etc.) and maintenance efficiency is connected to direct maintenance cost (cost for maintenance labor etc.) [9]. These two must be linked and analyzed in any improvement initiative. Exemplified: increasing maintenance effectiveness without taking into account maintenance efficiency can render in that the direct maintenance cost increases more than what is saved in decreasing indirect cost. Has an improvement really been implemented or has the improvement initiative been wasteful? (see further: [2]).

M. Bengtsson (✉)
Mälardalen University, Volvo Construction Equipment, Box 883, 721 23 Västerås, Sweden
e-mail: marcus.bengtsson@mdh.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_23

Using Pareto analysis (commonly known as the 80/20-rule or the rule of the "vital few and trivial many" [4]) in prioritizing which improvement initiative to undertake is one practical option and has been advocated for, as one tool (of many), to use within maintenance by maintenance research practitioners [3, 6–8, 10–12, 15]. For instance, [6, 7] uses Pareto analysis in order to assess the worst performing machines with the criteria downtime and frequency of calls. Duffuaa and Ben-Daya [3] state that Pareto analyses and diagrams can be utilized to identify: factors impairing productivity, crafts causing major backlogs, spare parts causing most delays, the most costly spare parts, and equipment causing the longest downtime. These types of visualization can be utilized when, for instance, prioritizing which equipment should be chosen for improvement work. However, there are some drawbacks with Pareto analysis if one is not cautious.

Sanders [13], for instance, makes a point in that a Pareto diagram, showing the distribution of any data, is like a photograph and that it tells nothing about the past or the future. Further [13] states: "Concentrate on the 20 % that are important today, of course. However, one can't disregard the 80 %, some of which may be found among the important 20 % tomorrow" (p. 40). When it comes to identifying failure codes that represents the majority of maintenance cost or downtime, [5] points out three possible deficiencies with Pareto analysis. (1) It can be difficult to find out which factor, or factors, that are dominant in contributing to cost or downtime if the Pareto histogram is based on downtime or cost alone. (2) With Pareto analysis one might miss identifying sporadic events with high maintenance cost or long downtime, or frequent events that individually does not cause high maintenance cost or long downtime but in total consumes much resource. (3) Pareto analysis is not generally applicable for trending purposes. As such, it is necessary to treat Pareto analysis with some precautions when prioritizing maintenance improvement initiatives.

The main objective of this research is to develop and test a tool to be utilized for prioritization purposes of maintenance improvement initiatives. The purpose of the paper is to exemplify, from an industrial case, some of the strengths and weaknesses of this approach.

2 Case Study Context

In this study, data from three factors from a computerized maintenance management system (CMMS) from one large production site in Sweden have been studied and analyzed in three individual Pareto analyses and further been developed into a tool to be used in prioritizing between different maintenance improvement initiatives. The production site manufactures, assembles, and paints components for the automotive industry. Roughly 700 employees tend roughly 300 manufacturing machines, various assembly equipment, test benches, a hardening shop, and a paint shop. CMMS-data from 2013 was downloaded. In total, any type of maintenance activities (i.e., immediate and/or deferred corrective maintenance and/or predetermined or condition based maintenance (for definitions please consult: SS-EN 13306 [14])) were performed on 1180 machine objects in the production site. In total, 516 machine objects suffered from breakdowns (immediate corrective maintenance). The maintenance department consists of three sub-departments: production maintenance, maintenance support, and maintenance engineering. At the sub-department production maintenance, four team leaders, with four to five electro-mechanics/repairmen per team, are responsible for their own geographical area of the production site. Production maintenance works daytime. Work tasks, such as: planning and execution of preventive maintenance, planning and execution of other planned work, improvement work, are included. At the sub-department of maintenance support, one team leader, with 20 electro-mechanics/repairmen, are responsible to support production maintenance in work peaks and are called out to repair breakdowns on machine objects. Maintenance support works in three shifts. At the sub-department of maintenance engineering, two team leaders (with ten maintenance engineers per team) works with support to the other maintenance sub-departments as well as to other departments within the production site with tasks, such as: specialists function, education, condition monitoring, remanufacturing, supplier contact, maintenance management development etc. Maintenance engineering works daytime.

3 Initial Pareto Analyses

In analyzing the factor number of breakdowns, the 20 % of the 1180 machine objects with most breakdowns accounted for 87 % of the total number of breakdowns, see Fig. 1. In only analyzing the machine objects that had suffered from breakdowns during 2013, which amounted to 516 machine objects, the 20 % with most breakdowns accounted for 60 % of the total number of breakdowns. Secondly, in analyzing the breakdown-related work orders through the factor down time, the 20 % of the work orders with the longest down time accounted for 78 % of the total down time, see Fig. 2. Thirdly, in analyzing the factor breakdown-related clocked working hours on the work orders, the 20 % of the work orders with the most clocked working hours accounted for 66 % of the total amount of clocked working hours due to breakdowns, see Fig. 3.

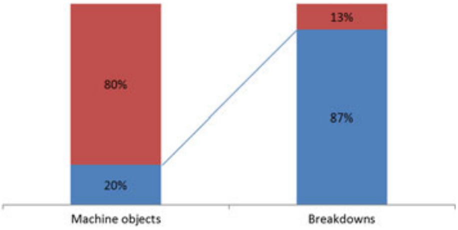


Fig. 1 Pareto analysis of number of breakdowns in machine objects. As visualized, the 20 % worst performing machine objects account for 87 % of the total number of breakdowns

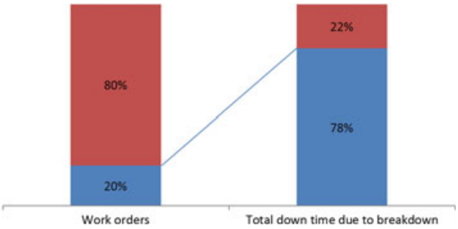


Fig. 2 Pareto analysis of the down time of work orders related to breakdowns. As visualized, the 20 % work orders with the longest down time account for 78 % of the total down time related to breakdowns

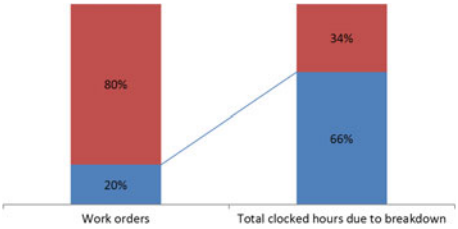


Fig. 3 Pareto analysis of clocked working hours of work orders related to breakdowns. As visualized, 20 % of the work orders with the most clocked working hours account for 66 % of the total clocked working hours related to breakdowns

The analyses were found to be interesting individually but questions were raised as to how they should be used in prioritization purposes. In using all of them, individually, many machine objects and work orders would be prioritized. Using only one of the analyses could imply a risk in that the machine object with most influence on the total sum might be lost.

4 Multi-factorial Pareto Analysis

4.1 Development of the Tool

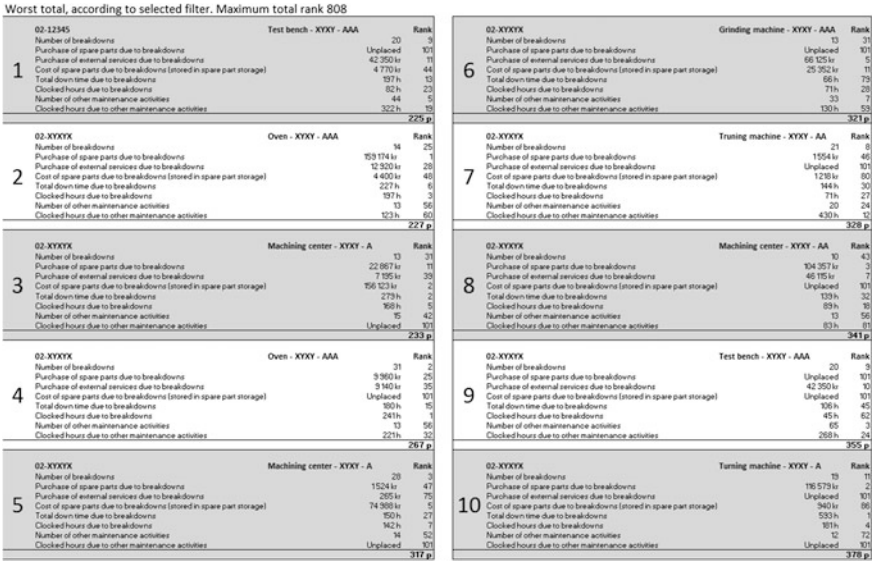
The solution was instead to give ranking order of the machine objects in all respective Pareto analyses (factors) to be added up to one total ranking number per machine object. The ranking orders, per Pareto (factor), were between 1 and 100; where 1 should be considered the “worst” machine object in respective Pareto analysis (factor) (all objects which were not on the worst 100 list were given the rank 101). In order to get an even wider view of influences on machine objects, additional data from in total eight factors, from the CMMS, were added:

1. number of breakdowns,
2. total down time due to breakdowns,

3. total clocked hours due to breakdowns,
4. total cost of purchased external services due to breakdowns,
5. total cost of purchased spare parts due to breakdowns (not stored in spare part storage),
6. total cost of spare parts due to breakdowns (stored in spare part storage),
7. number of other maintenance activities (disturbances and other maintenance work), and
8. total clocked hours due to other maintenance activities.

Therefore, the lowest possible total ranking would be 8 (worst possible in all Pareto) and the highest 808 (as several machine objects can score 101 in respective Pareto analysis several machine objects can have a total ranking of 808). The data are extracted monthly from the CMMS and entered into an Excel-sheet where the total ranking per machine object is calculated, see Figs. 4 and 5. The Excel-sheet with the total ranking are created in two versions, one contains data from 6 month history and the other 3 month history.

Further, various filtering options were entered into the Excel-sheet in order to give potential users the possibility to set their own filters for what they might find to be interesting, see Fig. 6. For instance, production managers have the possibility to choose only to view his/her own departments to see which machine object are suffering most from maintenance problems. Similarly, maintenance managers and



Worst total, according to selected filter. Maximum total rank 808

1

02-12345	Test bench - XYXY - AAA	Rank
Number of breakdowns	20	9
Purchase of spare parts due to breakdowns	Unplaced	101
Purchase of external services due to breakdowns	42 350 kr	11
Cost of spare parts due to breakdowns (stored in spare part storage)	4 770 kr	44
Total down time due to breakdowns	197 h	13
Clocked hours due to breakdowns	82 h	23
Number of other maintenance activities	44	5
Clocked hours due to other maintenance activities	322 h	19
		225 p

2

02-XYXYX	Oven - XYXY - AAA	Rank
Number of breakdowns	14	25
Purchase of spare parts due to breakdowns	159 174 kr	1
Purchase of external services due to breakdowns	12 920 kr	28
Cost of spare parts due to breakdowns (stored in spare part storage)	4 400 kr	48
Total down time due to breakdowns	227 h	6
Clocked hours due to breakdowns	197 h	3
Number of other maintenance activities	13	56
Clocked hours due to other maintenance activities	123 h	60
		227 p

3

02-XYXYX	Machining center - XYXY - A	Rank
Number of breakdowns	13	31
Purchase of spare parts due to breakdowns	22 867 kr	11
Purchase of external services due to breakdowns	7 195 kr	39
Cost of spare parts due to breakdowns (stored in spare part storage)	156 123 kr	2
Total down time due to breakdowns	273 h	2
Clocked hours due to breakdowns	168 h	5
Number of other maintenance activities	15	42
Clocked hours due to other maintenance activities	Unplaced	101
		233 p

4

02-XYXYX	Oven - XYXY - AAA	Rank
Number of breakdowns	31	2
Purchase of spare parts due to breakdowns	9 960 kr	25
Purchase of external services due to breakdowns	9 140 kr	35
Cost of spare parts due to breakdowns (stored in spare part storage)	Unplaced	101
Total down time due to breakdowns	180 h	15
Clocked hours due to breakdowns	241 h	1
Number of other maintenance activities	13	56
Clocked hours due to other maintenance activities	221 h	32
		267 p

5

02-XYXYX	Machining center - XYXY - A	Rank
Number of breakdowns	28	3
Purchase of spare parts due to breakdowns	1524 kr	47
Purchase of external services due to breakdowns	265 kr	75
Cost of spare parts due to breakdowns (stored in spare part storage)	74 988 kr	5
Total down time due to breakdowns	150 h	27
Clocked hours due to breakdowns	142 h	7
Number of other maintenance activities	14	52
Clocked hours due to other maintenance activities	Unplaced	101
		317 p

6

02-XYXYX	Grinding machine - XYXY - AAA	Rank
Number of breakdowns	13	31
Purchase of spare parts due to breakdowns	Unplaced	101
Purchase of external services due to breakdowns	66 125 kr	5
Cost of spare parts due to breakdowns (stored in spare part storage)	25 352 kr	11
Total down time due to breakdowns	66 h	79
Clocked hours due to breakdowns	71 h	28
Number of other maintenance activities	33	7
Clocked hours due to other maintenance activities	130 h	59
		321 p

7

02-XYXYX	Truning machine - XYXY - AA	Rank
Number of breakdowns	21	8
Purchase of spare parts due to breakdowns	1554 kr	46
Purchase of external services due to breakdowns	Unplaced	101
Cost of spare parts due to breakdowns (stored in spare part storage)	1 218 kr	80
Total down time due to breakdowns	144 h	30
Clocked hours due to breakdowns	71 h	27
Number of other maintenance activities	20	24
Clocked hours due to other maintenance activities	430 h	12
		328 p

8

02-XYXYX	Machining center - XYXY - AA	Rank
Number of breakdowns	10	43
Purchase of spare parts due to breakdowns	104 357 kr	3
Purchase of external services due to breakdowns	46 115 kr	7
Cost of spare parts due to breakdowns (stored in spare part storage)	Unplaced	101
Total down time due to breakdowns	139 h	32
Clocked hours due to breakdowns	89 h	18
Number of other maintenance activities	13	56
Clocked hours due to other maintenance activities	83 h	81
		341 p

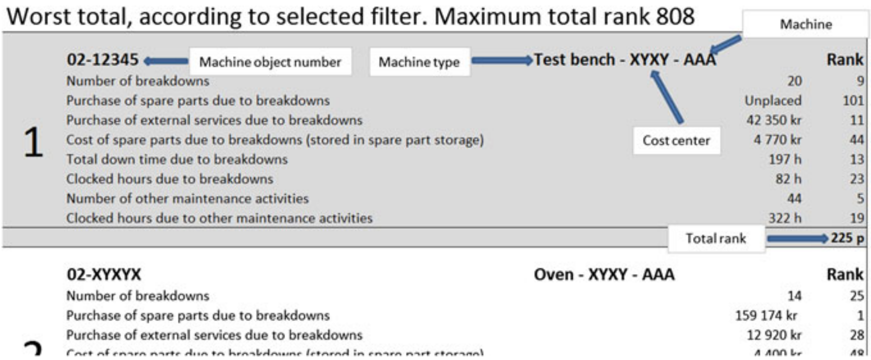
9

02-XYXYX	Test bench - XYXY - AAA	Rank
Number of breakdowns	20	9
Purchase of spare parts due to breakdowns	Unplaced	101
Purchase of external services due to breakdowns	42 350 kr	10
Cost of spare parts due to breakdowns (stored in spare part storage)	Unplaced	101
Total down time due to breakdowns	106 h	45
Clocked hours due to breakdowns	45 h	62
Number of other maintenance activities	65	3
Clocked hours due to other maintenance activities	268 h	24
		355 p

10

02-XYXYX	Turning machine - XYXY - A	Rank
Number of breakdowns	19	11
Purchase of spare parts due to breakdowns	116 579 kr	2
Purchase of external services due to breakdowns	Unplaced	101
Cost of spare parts due to breakdowns (stored in spare part storage)	940 kr	86
Total down time due to breakdowns	593 h	1
Clocked hours due to breakdowns	181 h	4
Number of other maintenance activities	12	72
Clocked hours due to other maintenance activities	Unplaced	101
		378 p

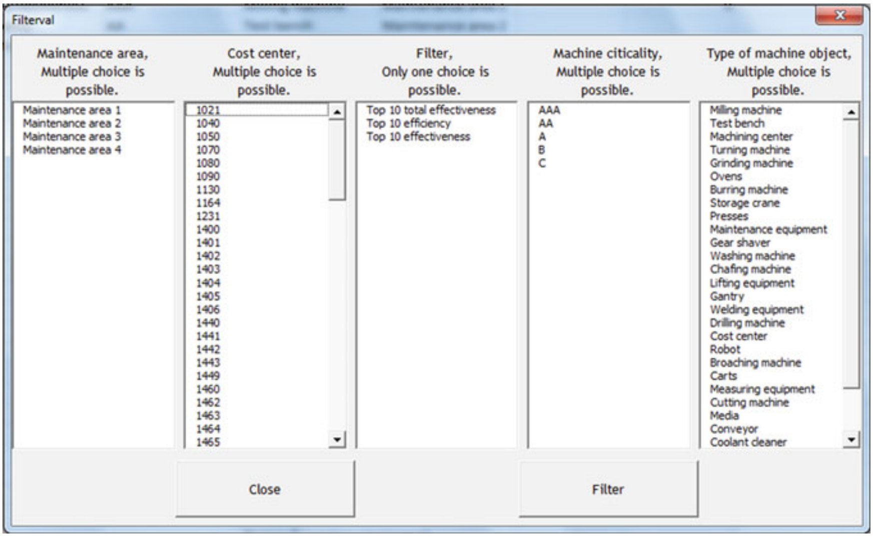
Fig. 4 Screen dump of one view in the Excel-sheet visualizing the top 10 worst machine objects (between February and July of 2014) when filtering on all possible machine objects in the site as well as all possible Pareto (factors). Machine object numbers, machine types, and cost center number are fabricated. See Fig. 5 for close-up



Worst total, according to selected filter. Maximum total rank 808

	02-12345	Test bench - XYXY - AAA		Rank
1	Number of breakdowns		20	9
	Purchase of spare parts due to breakdowns		Unplaced	101
	Purchase of external services due to breakdowns		42 350 kr	11
	Cost of spare parts due to breakdowns (stored in spare part storage)		4 770 kr	44
	Total down time due to breakdowns		197 h	13
	Clocked hours due to breakdowns		82 h	23
	Number of other maintenance activities		44	5
	Clocked hours due to other maintenance activities		322 h	19
				225 p
	02-XYXYX	Oven - XYXY - AAA		Rank
2	Number of breakdowns		14	25
	Purchase of spare parts due to breakdowns		159 174 kr	1
	Purchase of external services due to breakdowns		12 920 kr	28
	Cost of spare parts due to breakdowns (stored in spare part storage)		4 400 kr	48

Fig. 5 Close-up of the screen dump of the Excel-sheet. Some facts in the screen dump is highlighted and explained. Machine object numbers, machine types, and cost center number are fabricated



Filterval

Maintenance area, Multiple choice is possible.

Maintenance area 1
Maintenance area 2
Maintenance area 3
Maintenance area 4

Cost center, Multiple choice is possible.

1021
1040
1050
1070
1080
1090
1130
1164
1231
1400
1401
1402
1403
1404
1405
1406
1440
1441
1442
1443
1449
1460
1462
1463
1464
1465

Filter, Only one choice is possible.

Top 10 total effectiveness
Top 10 efficiency
Top 10 effectiveness

Machine citicality, Multiple choice is possible.

AAA
AA
A
B
C

Type of machine object, Multiple choice is possible.

Milling machine
Test bench
Machining center
Turning machine
Grinding machine
Ovens
Burring machine
Storage crane
Presses
Maintenance equipment
Gear shaver
Washing machine
Chafing machine
Lifting equipment
Gantry
Welding equipment
Drilling machine
Cost center
Robot
Broaching machine
Carts
Measuring equipment
Cutting machine
Media
Conveyor
Coolant cleaner

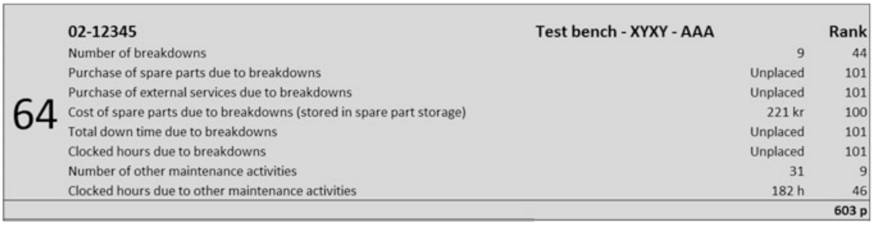
Fig. 6 A macro in the Excel-sheet makes filtering options easy

maintenance team leaders can filter different maintenance areas. Also, it is possible to filter according to machine criticality (AAA, AA, A, B, and C), for further information on machine criticality see [1]. It is possible to filter different type of objects (e.g., turning machines, grinding machines, robots etc.) which can be interesting for, e.g., production managers for comparison of similar machine object types located in different departments. Lastly, there is also a filtering option to either view all of the total ranking numbers or the ranking numbers with most influence on either the maintenance effectiveness or the maintenance efficiency. For the maintenance effectiveness the factors of: number of breakdowns, total time due to breakdowns, total clocked hours due to breakdowns, total cost of purchased spare parts due to breakdowns (not stored in spare part storage), total number of other maintenance activities, and total clocked hours due to other maintenance activities, were taken into consideration. These factors were judged by the maintenance management team to be causing most disturbances and variation for the customers (the production department). These factors include, for example, frequency and down time of maintenance work which directly impacts production. Also, the factor on total cost of purchased spare parts due to breakdowns (not stored in spare part storage) is included as it incurs longer waiting times as well as (often) a higher cost for shipping. This ranking can be utilized in order to see where maintenance improvement initiatives can have a good impact on customer satisfaction. When it comes to maintenance efficiency, the factors of: total clocked hours due to breakdowns, total cost of purchased external services due to breakdowns, total cost of purchased spare parts due to breakdowns (not stored in spare part storage), total cost of spare parts due to breakdowns (stored in spare parts), and total clocked hours due to other maintenance activities, were taken into consideration. These factors were judged by the maintenance management team to be causing most disturbances and variation within the maintenance organization. Clocked hours on work, both on breakdowns as well as other planned maintenance activities can be derived to several causes, such as for example, competence, training, and maintainability which all have impact on maintenance efficiency. Also, the factor on total cost of purchased external services due to breakdowns can be derived to competence issues but also lack of manpower or other planning issues. The two factors on spare parts concern cost-effectiveness of maintenance operations.

4.2 Utilization of the Tool—Maintenance Teams

The Excel-sheet, containing 6 months of history, with its Pareto analyses and total ranking number have been utilized by maintenance improvement teams in order to prioritize machine objects for maintenance improvement work. The total ranking number has been used as a start of the prioritization but additional input, such as communication with the customer (the production department) has also been taken into consideration in order to validate that improvements are truly needed, at least when it comes to improving maintenance effectiveness. As the Excel-sheet with 6 month of historic data is rather slow to use for follow-up of improvement work the 3 month version are used more often to validate that improvements have been successful.

As an example, one of the maintenance team leaders, with team members, in the sub-department production maintenance chose to work with improvements on the machine object number 1 (02-12345, fabricated number), visualized in Figs. 4 and 5, i.e., the worst machine object of all between February and July (of 2014). The goal of the improvement work was to eliminate three reoccurring failures and to



	02-12345	Test bench - XYXY - AAA		Rank
64	Number of breakdowns		9	44
	Purchase of spare parts due to breakdowns		Unplaced	101
	Purchase of external services due to breakdowns		Unplaced	101
	Cost of spare parts due to breakdowns (stored in spare part storage)		221 kr	100
	Total down time due to breakdowns		Unplaced	101
	Clocked hours due to breakdowns		Unplaced	101
	Number of other maintenance activities		31	9
	Clocked hours due to other maintenance activities		182 h	46
				603 p

Fig. 7 A screen dump visualizing the same machine as in Fig. 5 (between July and December of 2014). The improvement work on the machine ran between June and October. As visualized, the machine object, which were the worst possible (see Fig. 5) is now at 64th place. Machine object number, machine type, and cost center number are fabricated

improve availability generally. The improvement work ran between June and October (of 2014) while many activities were performed (it is out of the scope of this paper to present the improvement work methods). The results of the improvement activities were followed-up continuously and specifically the three reoccurring failures that were set out to be eliminated.

However, to visualize results of the improvement work, primarily for the customer (production manager and operators), on a general level, revisiting the Excel-sheet a few months after the improvement work has ended can tell how successful the improvement work have been. See Fig. 7 for data between July and December (of 2014). First, make a note that the machine object, which was the worst of all in February to July (of 2014), is now the 64th worst on the list. Also, make a note that seven factors have been improved (the eighth factor was, in both analysis, outside of the top 100 list). In this example, only the filter with all factors is being visualized, e.g., the maintenance effectiveness and maintenance efficiency filters are not visualized. However, it is rather clear that both maintenance effectiveness (fewer breakdowns, less total downtime etc.) and maintenance efficiency (less purchase of external services, fewer clocked working hours etc.) has been improved. The effectiveness and efficiency could of course be improved even more, particularly in the factor number of other maintenance activities which is still ranked as one of the worst in the production site.

4.3 Utilization of the Tool—Maintenance Management Team

Another possibility in the Excel-sheet is to view individual Pareto, on the different factors. This view can be utilized to zoom into potential structural problems in the respective factors, see further Figs. 8, 9, 10, 11, 12, 13, 14 and 15. The individual

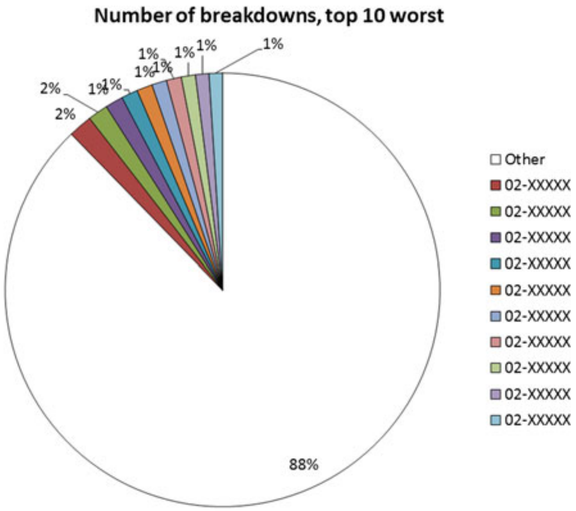


Fig. 8 Screen dump of one view in the Excel-sheet visualizing that the 10 worst machine objects (between June and November) in respect to number of breakdowns accounts for 12 % of all breakdowns

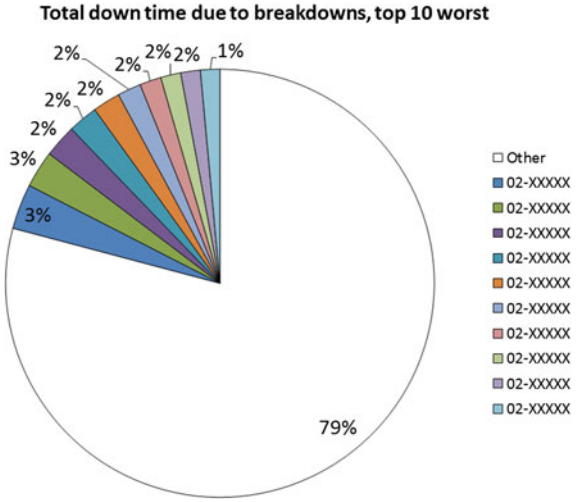


Fig. 9 Screen dump of one view in the Excel-sheet visualizing that the 10 worst machine objects (between June and November) in respect to total down time due to breakdowns accounts for 21 % of the total down time due to breakdowns

Pareto has been visualized only by the ten worst machine objects in respective factor in order to increase awareness on how much impact very few machine objects have. Also, these views are an important tool for the maintenance management team in order for them to work with diagnostics and improvement work in, for example, reducing the amount of external services and in optimizing the spare part storage.

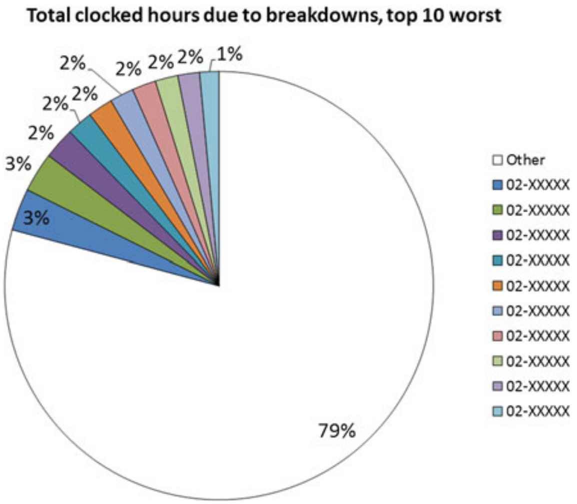


Fig. 10 Screen dump of one view in the Excel-sheet visualizing that the 10 worst machine objects (between June and November) in respect to total clocked hours due to breakdowns accounts for 21 % of all clocked hours due to breakdowns

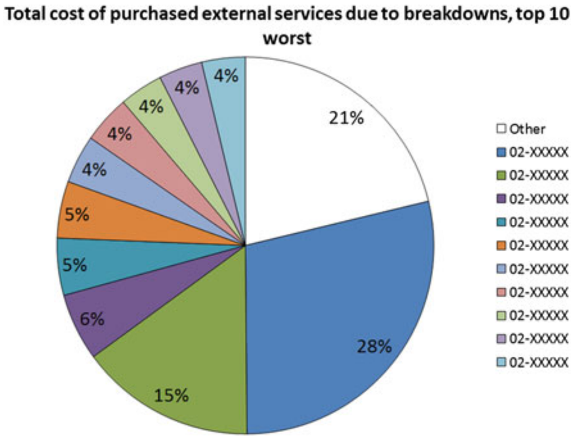


Fig. 11 Screen dump of one view in the Excel-sheet visualizing that the 10 worst machine objects (between June and November) in respect to total cost of purchased external services due to breakdowns accounts for 79 % of the total cost of purchased external services due to breakdowns

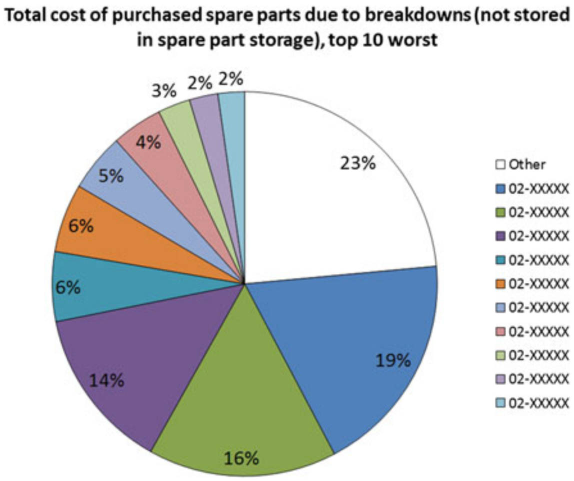


Fig. 12 Screen dump of one view in the Excel-sheet visualizing that the 10 worst machine objects (between June and November) in respect to total cost of purchased spare parts due to breakdowns (not stored in spare part storage) accounts for 77 % of the total cost of purchase of spare parts (not stored in spare part storage)

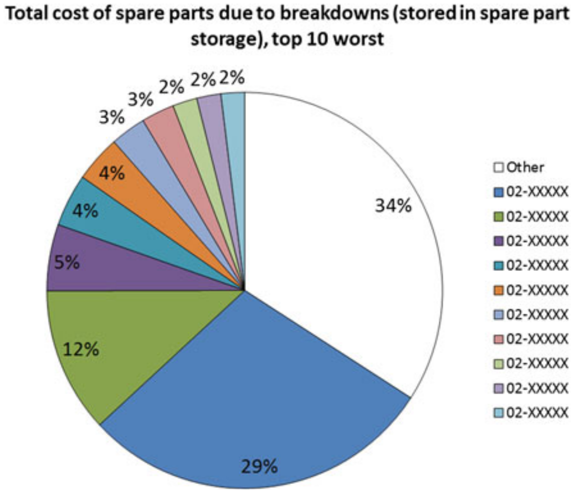


Fig. 13 Screen dump of one view in the Excel-sheet visualizing that the 10 worst machine objects (between June and November) in respect to total cost of spare parts due to breakdowns (stored in spare part storage) accounts for 66 % of the total cost of spare parts due to breakdowns (stored in spare part storage)

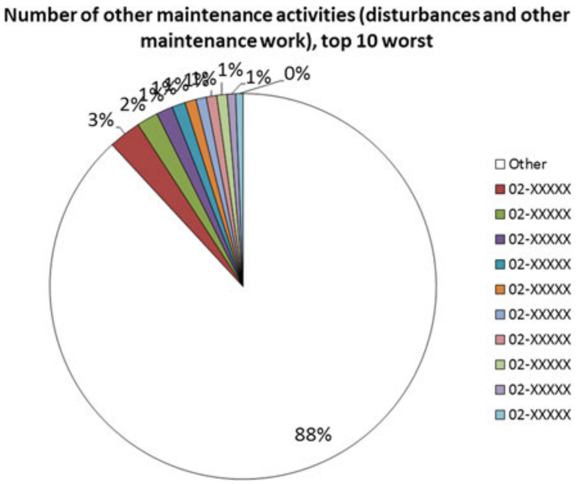


Fig. 14 Screen dump of one view in the Excel-sheet visualizing that the 10 worst machine objects (between June and November) in respect to number of other maintenance activities accounts for 12 % of all other maintenance activities

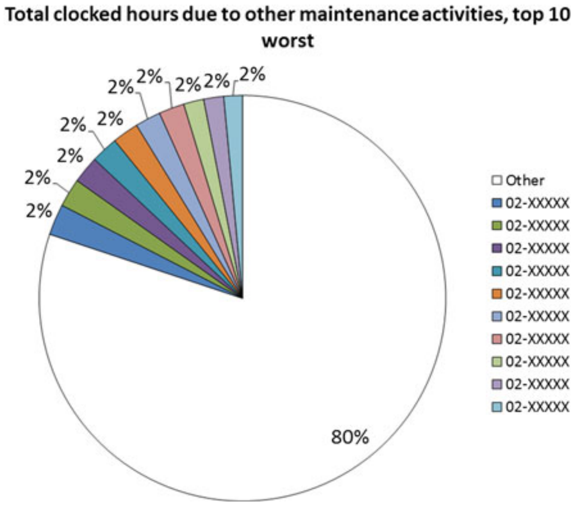


Fig. 15 Screen dump of one view in the Excel-sheet visualizing that the 10 worst machine objects (between June and November) in respect to total clocked hours due to other maintenance activities accounts for 20 % of all clocked hours due to other maintenance activities

5 Discussion and Conclusions

A CMMS being used in a large production site with a reasonable amount of machine objects quickly fills up with data. This data can and should in many cases be considered as experience. Making use of this data and experience is important in order to become more effective and efficient. It can, though, be a difficult task in finding useful and targeted information when, for example, trying to prioritize between different improvement initiatives. This research started with Pareto analyses of three individual factors from a historic base of 1 year from a CMMS and was developed into a tool taking into consideration and combining eight factors into one overall ranking number on machine object level. Future development of the tool includes weighting of the different factors. Some can be considered more influential than other, particular when splitting the total effectiveness into maintenance effectiveness and efficiency. For example, the number of breakdowns and maintenance activities in conjunction with total downtime is most likely more influential than perhaps cost of spare parts when it comes to maintenance effectiveness etc.

It is important that the tool is to be used as a tool to verify the customers' experiences rather than as a tool to strictly prioritize from. The need of the customer must always be in focus. The true need of an improvement is not always visible in a CMMS, other ways, for example, through human communication or OEE-measurements must be used in order to truly validate the internal need and future internal value of an improvement [2].

Acknowledgments This research work has been funded by VINNOVA through the "FFI—Hållbar produktionsteknik" research programme, and Mälardalen University. The research work is also a part of the initiative for Excellence in Production Research (XPRES) which is a cooperation between Mälardalen University, the Royal Institute of Technology, and Swerea. XPRES is one of two governmentally funded Swedish strategic initiatives for research excellence within Production Engineering.

References

1. Bengtsson M (2011) Classification of machine equipment. In: Proceedings of 1st conference on maintenance performance measurement & management
2. Bengtsson M, Osterman C (2014) Improvements in vain—the 9th waste. In: Proceedings of Swedish production symposium, SPS14
3. Duffuaa SO, Ben-Daya M (1995) Improving maintenance quality using SPC tools. J Qual Maint Eng 1(2):25–33
4. Juran JM (1989) Juran on leadership for quality: an executive handbook. The Free Press, New York
5. Knights PF (2001) Rethinking Pareto analysis: maintenance applications of logarithmic scatterplots. J Qual Maint Eng 7(4):252–263
6. Labib AW (1998) World class maintenance using a computerized maintenance management system. J Qual Maint Eng 4(1):66–75
7. Labib AW (2004) A decision analysis model for maintenance policy selection using a CMMS. J Qual Maint Eng 10(3):191–202
8. Madu CN (2000) Competing through maintenance strategies. Int J Qual Reliab Manage 17 (9):937–949
9. Márquez AC, de León PM, Fernández JFG, Márquez CP, Campos MJ (2009) The maintenance management framework—a practical view to maintenance management. J Qual Maint Eng 15(2):167–178
10. Mjema EAM (2002) An analysis of personnel capacity requirement in the maintenance department by using a simulation method. J Qual Maint Eng 8(3):253–273
11. Mokashi AJ, Wang J, Vermar AK (2002) A study of reliability-centred maintenance in maritime operations. Mar Policy 26(5):325–335

12. Moubray J (1997) Reliability-centered maintenance—RCM II, 2nd edn. Industrial Press, Inc., New York
13. Sanders R (1992) The Pareto principle: its use and abuse. J Prod Brand Manage 1(2):37–40
14. SIS, Swedish Standards Institute (2001) Maintenance terminology. SS-EN 13306, SIS Förlag AB
15. Smith R, Hawkins B (2004) Lean maintenance—reduce costs, improve quality, and increase market share. Butterworth-Heinemann, Burlington

Predictive Modelling for Estimation of Railway Track Degradation

Bjarne Bergquist and Peter Söderholm

Abstract The degradation processes affecting railway track condition depends both on the resistance of the track and on the stresses subjected to it. Regarding the stresses, both their magnitudes and cycles are of importance when considering the degradation. Furthermore, the stresses have some regularity and variability in the time domain, while the degradation resistance of a track has some spatial regularity as well as variability. In addition, the condition measurements of track may be both irregular and contain measurement errors. Hence, it is challenging to model the condition of track to enable predictions and condition-based maintenance. However, wear prediction models could help to change large parts of the maintenance practice from predominantly corrective to preventive if both the deterministic and the stochastic components of the wear process can be estimated with sufficient accuracy. In this study, one-step-ahead predictions have been used for establishing prognostic models based on repeated measurements of railway track geometry to estimate track wear properties, degradation rates and stochastic behaviour including measurement errors. The prognostic models have then been used for condition assessment and state predictions. Repeated sampling allows for estimations of measurement errors, but the irregular sampling need to be accounted for by interpolation in the time series modelling approach.

Keywords Spatiotemporal analysis · Condition-based maintenance · Prediction · Prognostics · Interpolation methods · Railway · Track · Degradation · Statistical Process Control (SPC) · Control charts

B. Bergquist (✉)
Luleå University of Technology, 971 78 Luleå, Sweden
e-mail: bjarne@ltu.se

P. Söderholm
Trafikverket (Swedish Transport Administration),
Box 809, 971 25 Luleå, Sweden
e-mail: peter.soderholm@trafikverket.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_24

1 Introduction

Condition monitoring of railway tracks is often performed by special measurement trains, including wagons that measure properties of the track, the overhead wire and the track substructure. The measurements are obtained at intervals that are influenced by the magnitude of the load subjected to the railway track such as frequency of trains and their axle load. Repeated measurements of railway track properties suggest that the data could be used to generate models for prediction of the railway track degradation. With reliable prediction methods, both the infrastructure manager and the entrepreneur would be given a means for longer planning horizons, enabling more preventive and less corrective maintenance. Changes from corrective maintenance to preventive maintenance would not only be more cost effective, but also reduce train delays and accident risks. The cost effectiveness is improved since longer planning horizons would increase the possibilities for effective use of expensive equipment and personnel and thus reduce overall maintenance costs. The delays and safety are improved since the maintenance proactively manages failure events before they develop into faulty states, by a surveillance of the changing condition of the track.

1.1 Statistical Process Control

Statistical Process Control (SPC) is a classical statistical approach used for many surveillance purposes within various sectors [1]. SPC is based on control charts, where measurements are compared to control limits. These limits are based on the statistical distribution of the sampled data.

Within SPC, an observation is classified as being within its expected range if it remains within the control limits, and if not, the assumption is that the process is affected by systematic variation and needs attention. Commonly it is assumed that the sampled data is normally distributed and independent. Special control charting techniques have been designed to be used when these assumptions do not hold. Regularly, control limits equal to three standard deviations of the studied property variation is used, which for normally distributed and independent data that are unaffected by assignable causes for variation would generate false alarms in 1 out of 370 observations. The detection capability can be described similarly. If an assignable cause was to shift the mean value of the process by 1 standard deviation (σ), a regular so called individuals x-chart with control limits of 3σ would have a 2.3 % chance of detection already at the first observation of the process after the shift, and a 50 % chance of detection of the assignable cause generating a deviation as large as 3σ from the nominal value. See also [2].

Control charts are often based on sampling procedures such as that five items are sampled, and then the level of the output is studied by plotting the average ($\bar{x}$) in one chart, and the variation is studied by plotting the sample standard deviation (s) in another, the so called $\bar{x}$ -s charts. Changes in any of these properties will thus render alarms. When studying individual observations (x), these are also plotted in a control chart with the purpose of monitoring the level of the output, and then the variation between consecutive measurements (moving range, MR) are used to monitor the variation in another chart, the so called x-MR charts.

In some situations, both the location as well as the time could be important for monitoring a certain characteristic, that is, the monitoring scheme requires spatiotemporal information. Spatiotemporal data using two-dimensional information have been suggested for railway condition monitoring, see Bergquist and Söderholm [3, 4].

The spatiotemporal approach introduced by Bergquist and Söderholm [4] is illustrated in the control cart of Fig. 1. In the control chart, twist variation has been plotted for each passage of the measurement wagon for each of the thirteen 150 m sections based on the Z-chart. The top chart displays the double logarithm of the twist range ((ln(ln($range$($twist$)))), and the two-observation moving range (MR) of these observations is represented in the lower chart. The median moving range has been used to calculate the control limits for both charts, since the median is not affected by drastic changes. The variation increases over time for each of the 13 sections, and now and then the variation abruptly drops. These variation drops usually coincide with known maintenance actions, i.e. tamping. Now, although this control charting technique renders much faster alarms than the regular procedure of using geometric alarm limits [3], it does not take full advantage of the time

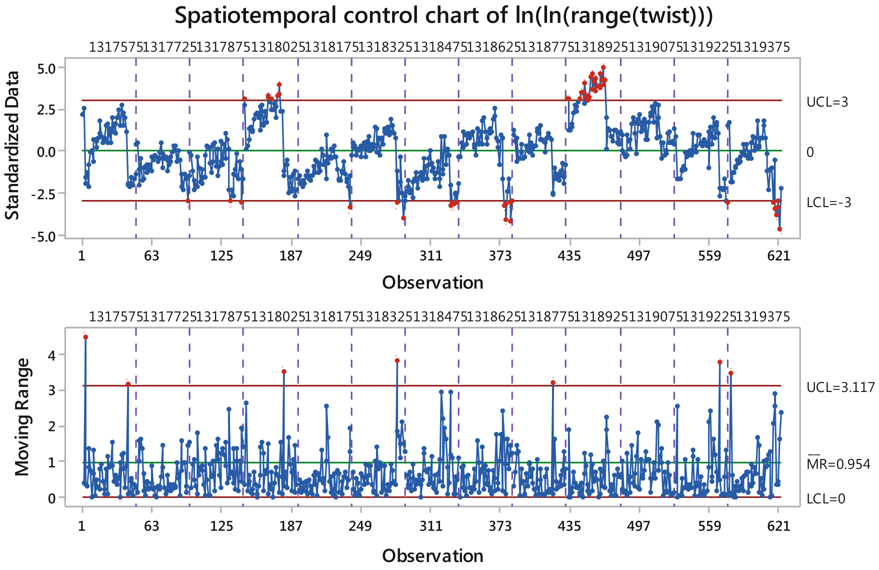


Fig. 1 Logarithm of range of twist of track divided into 150 m sections. Each observation represents one measurement occasion; the oldest to the leftmost of each section

dependence of the data. From Fig. 1, it is obvious that the condition is degrading with time, which means that the control limits will depend on long-term deviations.

In this study, we are interested in studying the dynamic evolvement of the track properties. This means that we are interested in alarms not just when the latest observations reach a certain limit, but also if the rate of deterioration (or improvement) would occur. One way to monitor the rate of change rather than the change would be to study the difference between repeated measurements directly as the primary response. The sequential observation differences are regularly monitored using the moving range chart that often accompanies the individuals control chart. In fact, the lower graph in Fig. 1 shows such a chart. The alarms of the moving-range chart seen in Fig. 1 are results of known and suspected maintenance actions. It is clear that the large, dramatic changes of the measured property will render alarms, but the alarm limits are too wide to detect smaller deviations. The proposed procedure in [3] does not render prognoses for the studied property either. By visually inspecting the graph, an idea of where the next observation could be expected is possible, but the irregular sampling procedure is not taken into account.

Here we suggest a more formal approach based on predicting the state of the studied property at a certain time by modelling the behaviour prediction process. We study interpolation methods as well as least squares estimation of polynomial equations, and compare the results by studying standard deviation of the prediction error.

Another difficulty generating a prediction model is acknowledging that the process is regularly maintained, and the deterioration of a property as well as the deterioration rate may not only be brought to a halt, but the state may be improved and the deterioration rate may be reduced as the result of such maintenance. A good prediction model should be able to handle such step changes. Since the condition of the first measurement following the maintenance action is likely to differ between different maintenance actions, making the prediction model accurate already for the first measurement after maintenance is likely difficult. A good model would, however, succeed to generate fair predictions for observations from the second measurement following the maintenance action. This paper intends to report on complementary approaches, where the deterioration rate itself is acknowledged, and observations that do not follow the expected deterioration process for a track section is scrutinized.

Time series analysis is a well-known method for the generation of future predictions and could therefore be a candidate for predictive maintenance actions, which in turn, includes promises for a number of positive outcomes. However, the sampling interval is also affected by climatic circumstances that may affect the track or the accuracy of the measurements such as spring thaw period, or logistic reasons such as operator vacancies and availability of measurement trains. Hence, where repeated measurement data are delivered at irregular intervals, the regular sampling assumption of times series analysis is violated. To overcome this challenge, this study aims to devise a prognosis method based on approaches that could be applied to the irregularly sampled data.

To fulfil the aim of the study, different prognostic models are compared with each other by usage of condition data measured at a track section of the Swedish Iron ore line, which is a heavy haul line in the northern part of Sweden.

1.2 Interpolation Techniques

Interpolation is used to calculate intermediate values and convert disjoint data points to a continuous function. The methods differ, for instance in the way the derivative is required to be continuous as well. Nearest neighbour interpolation simply uses the value of the nearest data point in between samples, while linear interpolation connects points through linear functions. Many spline methods, or Kriging methods [5], generate continuous derivatives, and thus create smoother interpolation functions with curves lacking sharp corners. The Kriging formula in this case generates an estimate, $\hat{Z}$, of the unmeasured property at a time t_0 between observations, according to Eq. 1.

$$\hat{Z}(t_0) = \sum_{i=1}^{N} \lambda_i Z(t_i) \tag{1}$$

Where: Z(t_i) is the measured property values at the *i*th time, λ_i is the Kriging weight constant, and N is the number of measured values to use for the interpolation.

Spline functions are regularly used for interpolation, but when data may contain noise, the regular spline functions, such as the cubic splines, tend to oscillate and be susceptible to outliers. A regular spline function has global propagation, and the whole spline function will be affected if there is an outlier anywhere among the measurements, regardless if the outlier was detected a long time ago. Splines with local propagation, meaning that the closest control points (measurements) have the largest importance for the curve near these, will improve fitting and are more promising when seeking an extrapolation model. Splines with local propagation include the Akima spline [6] and the B-spline (also known as the basis spline). The Akima interpolation spline is a continuously differentiable sub-spline that is piecewise, meaning that the nearest neighbours influence the interpolation values. The curve is therefore split into segments and each segment is influenced only by a defined set of nearest neighbours. The interpolation function is defined as in Eq. 2:

$$\hat{Z}(t) = k_0 + k_1(t - t_i) + k_1(t - t_i)^2 + k_1(t - t_i)^3, \ldots, t_i \le t \le t_{i+1} \tag{2}$$

where the constants are determined by the first derivatives t_i' and t_{i+1}' at the at the endpoints of the interval, see [6]. The Akima spline is compared to other spline types, such as cubic splines, more robust versus outliers.

For an overview of the result of five different interpolation methods, see Fig. 2.

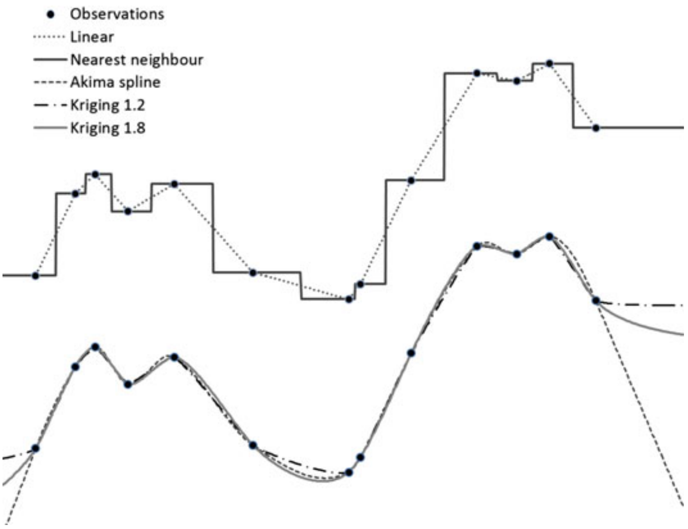


Fig. 2 Irregularly sampled observations and different interpolation methods. Kriging method constants in legend. As seen, these interpolation methods fit the curve to the observations while interpolation values differ

Forcing the interpolation methods to pass through the observation may not always be the best choice. The reason for this is that observations usually carry with them some amount of error, e.g. from the measurements.

2 Method

The overall study approach selected to fulfil the aim of this study was a single-case study of the Swedish iron ore line. Quantitative data representing the track condition was collected by measurement trains running along one selected track section of the line, e.g. due to the large number of runs, which supported the modelling efforts. The analysis of empirical data was mainly based on a combination of theories from the fields of SPC, time series analysis, and interpolation methods. This research method is described in some more detail in the following subsections.

2.1 Studied Case

Track condition data for the Swedish railway system are collected by measurement wagons that regularly are pulled along the track system in speeds up to 200 km/h. Each section of the Swedish track system is measured up to six times per year depending on the section's criticality. Observations of about 30 track geometry variables are obtained and stored for every 25th cm. Track variables include the position coordinates (height, and locations in the plane), and the track width [7, 8].

The collected data is then stored in a database (Optram [9, 10]) together with positioning and measurement time information. The database also contains information about the infrastructure and its attributes (e.g. type of object, geographical position, and description) and if the measurement is taken on a point asset (e.g. railway switch and level crossing) or a linear asset (e.g. track and catenary system). Other Optram information includes data about maintenance events and their history, e.g. track alignment and related information.

The empirical data used for this study was collected from Optram ranging from April 27, 2007 to November 21, 2014. The track was chosen as the major study object and a track section at the Swedish iron ore line was chosen based on that it did not contain any switches and crossings, platforms et cetera, nor any sharp curves or had had the track replaced during the chosen timeframe. In addition, derailment critical geometry faults had been found on the section, which also was a selection criteria, since one of the study goals was to see if such faults could be predicted.

The chosen 2 km track is found on track Section 113, 1327 km and 500 m to 1319 km and 450 m. This is the track connecting the mining towns of Gällivare and Kiruna, and the section is found 5 km from Gällivare train station, approximately 100 km north of the Arctic circle. The Iron ore line is a single track railway with the western endpoint in Narvik harbour, Norway, and with southbound connections to the rest of Sweden placed at Boden. This means that track 113 is on a reversing section and therefore many of the studied 48 measurements are only separated by 1 day or a weekend, such that one measurement is taken when the measurement train passes on the way to the Norwegian boarder, and then passes again one or a few days after at the return from Norway. Compared to the times between other measurements, these passages could be classified as repeated measurements.

2.2 *Studied Response*

The cant is one of the critical track geometry variables, typically expressed as the difference in elevation of the two rails, a quantity referred to as the superelevation. On a straight part of the track, the two rails should be level, i.e. the cant should be zero. On a curved part of the track, the cant denotes the raising of the outer rail with respect to the inner rail to allow for banking. This banking is in turn needed to compensate for the generated centrifugal forces. However, the train may derail if the cant changes too rapidly. The cant change rate is called twist, i.e. the rate of change of the track superelevation and is defined as the algebraic difference between two cants taken at a defined distance apart. The cant is usually expressed as a gradient between the two points of measurement, i.e. as a ratio (% or mm/m). Twist measurements is either taken simultaneously at a fixed distance, e.g. at a distance equivalent to the wheel-base, or is computed from consecutive measurements of cant. Normally, the twist is measured on a 6 m base, i.e. the cant measured at two points with 6 m distance.

The studied time interval contains 48 twist measurement occasions. The measurements were performed by measurement trains by the entrepreneur company Infranord's measurement wagons STRIX, IMV 100 N and IMV200. Besides measurement wagons and instruments, the measurement speeds, and thus the dynamic forces subjected to the track during measurements differed between measurements. In total, 14 out of the 48 measurement occasions were classified as repeated measurements for measurement system analysis purposes. The times to the following measurement occasions classified as replicate measurement occasions were selected with a maximum of 3 days separation from a previous measurement.

The sampled twist data include major positioning errors, and therefore the studied 2 km section was split into 150 m long subsections. The range between the maximum and minimum twist of each 150 m section was used as a suitable response to overcome the positioning difficulties, and the logarithm of the logarithm of the twist was chosen as an appropriate measure, see Bergquist and Söderholm [3, 4].

2.3 Measurement Error Calculation

The 13 different 150 m sections were assumed to be independent, and the differences between the results of repeated measurement occasions for each section of 150 m length were used to calculate the measurement error variation for that replicate. The total measurement error was then calculated through pooling the standard deviation of all the 14 replicate measurements.

2.4 Interpolation Method Comparison

The irregularity of the sampling seen in Fig. 3 necessitates interpolation to obtain the presumed range data at regular intervals. The measurements' spread over the studied interval is found in Fig. 3a, b. The interpolation interval was chosen to 3 months, and the chosen dates were March 31, June 30, September 30 and December 31.

The methods used for the interpolation included the Akima spline, nearest neighbour, linear interpolation, Kriging interpolation with constants 1.05 and 1.2, and also least squares regression using both linear and quadratic estimations, see Fig. 4. Note that the regression allowed for the fitted curve not to pass through the observations, something that the other methods did not.

A time interval was chosen to contain training data recorded between October 9th, 2007 and October 1, 2009 and this interval included 11 observations each on

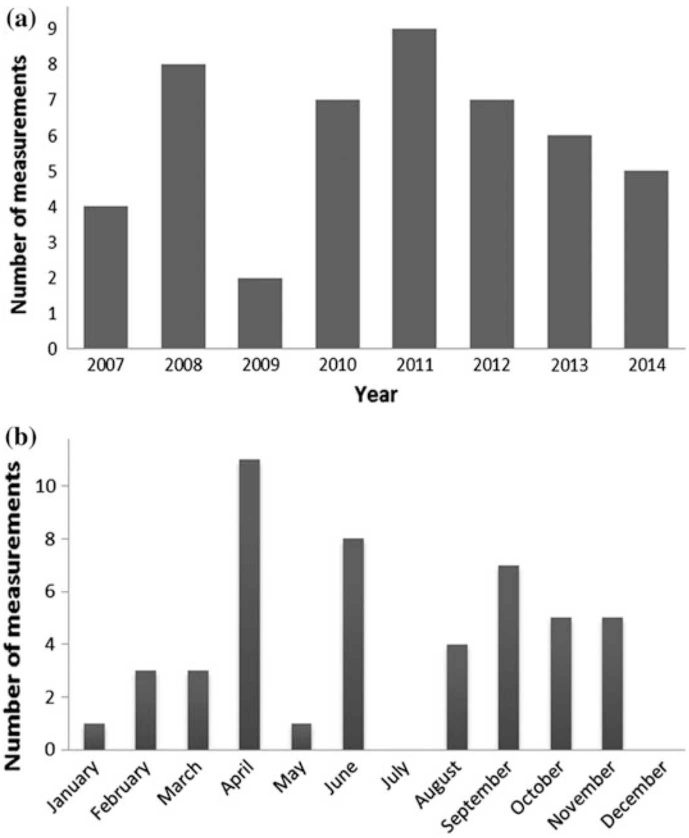


Fig. 3 **a** Number of measurements per year. **b** Number of measurements per months over the 8 years

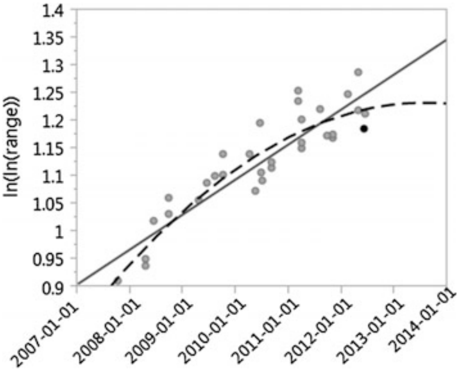


Fig. 4 Linear regression fitting (*solid grey curve*) and 2nd degree polynomial curve fitting (*black dashed curve*) to measurement data

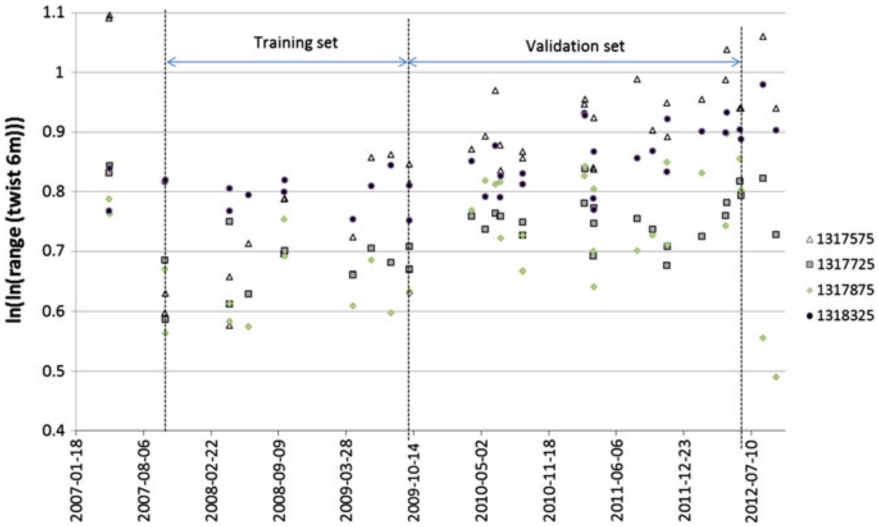


Fig. 5 Observations used for training the different interpolation methods and one-step-ahead extrapolation period used for validation

the 13 studied 150 m intervals. The interpolation methods were then tested by using a one-step-ahead prediction of the observation obtained during the validation period, ranging between October 2, 2009 and June 7, 2012. The intervals were chosen since neither 150 m section showed any dramatic changes of the observed property, indicating that no unreported maintenance actions had been performed. Data for four of the 13 sections obtained from the training period and the validation period are seen in Fig. 5.

After the one-step-ahead prediction and the prediction error was calculated by comparing the prediction with the measured value, the training set was expanded with the new observation, new models were calculated and new one-step-ahead predictions were calculated over the validation interval. The sum of the squared one-step-ahead prediction error was then used to evaluate which of the interpolation methods that gave best results. Note that the one-step-ahead prediction error is an extrapolation rather than an interpolation, and it is likely that methods that are sensitive towards the last observation (e.g. the Akima spline in Fig. 2), are ranked low using this procedure.

2.5 Predictive Control Charting Procedure

The suggested procedure uses an empirical model based on previous measurement for predictions for the state of the monitored variable, and whenever the database is updated, the prediction is compared to the observation, the residual is calculated and the model is updated using the new information.

Two methods to update the prediction model when the modelled data displays drastic changes as results of maintenance actions. Here, the twist variation along a given track section (measured as the twist range) is studied. The maintenance actions that will affect the twist variation for instance include tamping, track alignments, ballast change, sleeper changes, changing the track or combinations of the mentioned actions. All these can be expected to improve the twist and the twist variation conditions as these actions include alignments of the track. It is not unlikely that the actions also affect the degradation rate. As the maintenance action is expected to improve conditions for years to come whereas the properties are measured several times a year, step changes would be a simple way of modelling the maintenance effects on the response that may prove sufficiently accurate, and this approach is the one followed here.

Step changes may be implemented differently; we will study two approaches: In the first approach, we introduce the step to the old observations, so that the prediction model would fit with the new, improved state. The second approach entails using the same model but adjusting new predictions after the maintenance actions with the step change. Both approaches require estimation of the size of the step, and this step will likely differ for different applications. The size of step is here taken as equal to the magnitude of the first residual after the maintenance actions, that is the difference between the model prediction at this time and the actual observed value.

3 Results

In this section the results of the performed analysis are presented, i.e. the estimated measurement errors, the difference between compared interpolation methods, and finally the proposed control chart procedure.

3.1 Measurement Errors

Counting that each of the 14 replicate measurement contained information from 13 sub-sections, the pooled measurement error estimate was based on 182 replications. The pooled measurement standard deviation was found to be 31 E-3, and a 95 % confidence interval for an observation assuming normality and independence was calculated to be 0.12.

3.2 Interpolation Method Comparison

The sum of the squared prediction error, SQE, of the interpolation methods are given in Table 1. The Akima spline's sensitivity toward endpoint measurements

Table 1 One-step-ahead prediction error of studied methods

Method	**SQE**
Akima spline	294
Nearest neighbour	87.8E-3
Linear interpolation	87.8E-3
Kriging with constant 1.05	93.1E-3
Kriging with constant 1.2	122E-3
Linear regression	54.6E-3
2nd degree polynomial regression	58.7E-3

also makes it sensitive versus measurement errors, and therefore the measurement errors are large. The Kriging methods as well as using the nearest neighbour or using linear interpolations all give fair predictions. These methods are, however, beaten by the regression methods where the model does not pass through the observations. The more relaxed requirements of the whereabouts of the model near observations are reasonable given that the observations contain considerable measurement errors. The linear regression model and the regression model using cubic terms have similar prediction errors, but the linear regression is the more parsimonious, and is therefore chosen.

3.3 Control Charting Procedure for Spatiotemporal Data

As with regular control charting procedures, the intent of this study is also to signal when observations are found that deviate from the expected. However, we want to monitor a dynamic process that may have wandering means as well as variations following the deterioration of the track. The expectation thus is constantly changing as the track deteriorates from use or hopefully improves as a result of maintenance actions. Hence, unexpected improvements as well as unexpected reductions of the quality of the studied property are of interest. In the previous attempts [3, 4], reduced performance was detected only when passing conditions should be visible also for the first observation following the first one with large deviations, but models that do not consider corrective actions will continue to alert until the regression equation is based on sufficiently many late observations. Means to forget old observations include weighing the latest observations the most such as using the exponentially weighted moving average (EWMA) charts, but these would also compensate for worsening conditions and are based on a static mean.

Here, it is proposed that the control chart is based on the residual between the prognosis and the observation, and where the prognosis is based on estimation of an intercept and a time-based constant. The control chart primarily is a chart that controls the temporal model for the degradation or improvement of the track data. When the railway condition is maintained, the intercept of the temporal model of the studied property is updated by adjusting the old observations so that the curve

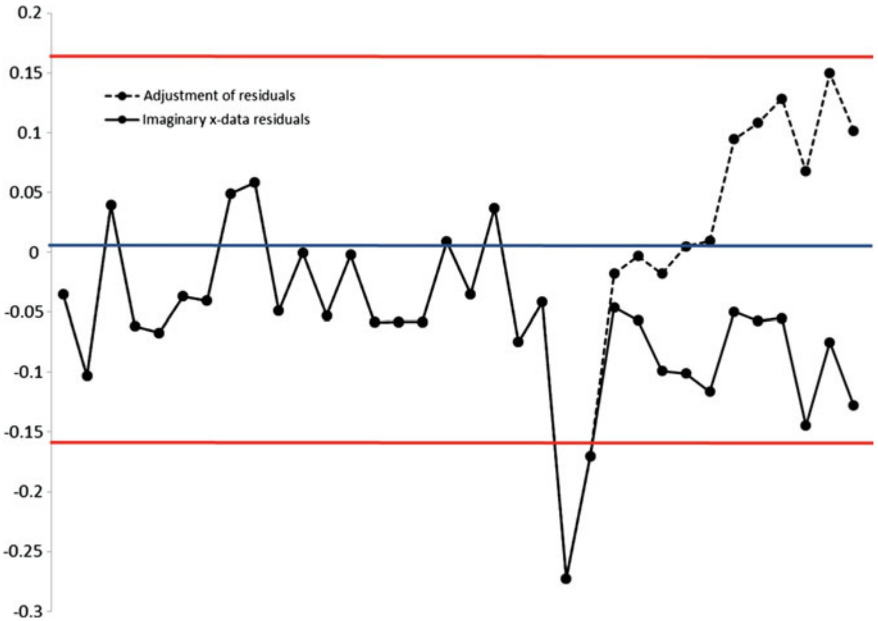


Fig. 6 Residuals control chart for track section 1318 Km, 850 to 1319 m 0, first approach where the prediction model for observations following the maintenance action is based on fictive prior observations

will pass through the new observations following the maintenance actions. In Fig. 6, the solid curve, such a control chart is depicted for the track section 1318 km, 850 to 1319 m 0. The residuals from the linear extrapolation prediction are plotted versus the measurement time. The chart signals for the August 6, 2012 observation, where the model had predicted a higher twist variation. Note that the signalling observation was not adjusted, since lack of information of maintenance actions are common in the Optram database, and in this case the deviation that would trigger an alarm.

The second approach, where instead the predictions are updated is also depicted in Fig. 6, i.e. the dashed curve. Note that the two methods generate two different paths, the former leading to increasingly negative residuals as new observations are added, the latter to increasingly larger positive residuals.

Now with the control chart, the old predictions before the alarms remain as is, but new predictions use the updated model for new predictions. It is not useful for the display of the prognostic model to show where the model was erroneous the same way it is for the control chart, so earlier erroneous predictions are adjusted when the control charts have signalled that the model is no longer generating correct predictions.

4 Discussion

The study showed that the spline method was much worse than the other methods, which only differed slightly in-between. The properties obtained by the measurement wagons all have errors stemming from the measurement. The measurement itself carries errors due to dynamic effects stemming from that successive measurements rely on measurement trains that in turn may travel at different speeds, or indeed be replaced so that the instrument differ between measurement. Another major source of error is that the positioning of the observation may differ more than 50 m. Spline methods forcing the model to pass through the observations when these contain considerable error resulted in oscillation of the model curve and also that the extrapolation direction was severely off target. The other methods passing through observations also performed worse than the regression that allowed the model to be influenced by the observation, but that did not slavishly follow the new observation.

For some sections, the 2nd order polynomial outperformed the linear model, and the more flexible polynomial model probably had been more advantageous for a response showing a more non-linear behaviour. Applying the logarithm twice on the twist range transforms a long-tailed distribution to one more closely resembling the normal distribution. Logarithm transformation is a standard transformation for variation that has a skew distribution. Logarithm transformations are also often applied to properties where the variation and level are tied to each other in a multiplicative manner, that is, when variation tend to increase as the average increases. Without the second logarithm applied to the twist range, the variation of the twist range would indeed have increased, so that the highest variation of the twist variation would be obtained for large twist ranges. As seen in Fig. 1, the variation pattern, although increasing, does seem constant when the slope is accounted for. The slope of the increase is close to constant for all sections; without the second logarithm transformation the slope instead would have increased between the maintenance events. For such responses, the 2nd degree polynomial is likely to outperform a regression model only containing a term for intercept and slope.

The calibration of the control charts for track alignment actions that were unaccounted for, such as the one that should have preceded the large negative residual of August 6, 2012 may need some consideration. That the track for some reason seems to be mended and straightened itself goes against the third law of thermodynamics and that is unlikely. A more probable cause would be related to the measurement systems. The repeated measurements did not reveal improvements after this point and other track sections showed a continued variation increase, but in a general case, a reduction of the range of the twist due to reduction in instrumental error is possible and the method readily allows for adoption and continuation of monitoring after such changes. Although such a track maintenance action was not reported in the database, we conclude that maintenance was the only reasonable explanation for this step reduction of the measured variation.

In this study, we investigated two procedures of adjusting the time series after maintenance actions. The first procedure was to make new prediction models based on a changed intercept due to tamping that deducted the values of all previous observations including that of the negative outlier that was generated from the maintenance action. Then the old model was recalculated, based on these imaginary observations, and this new model was then used for the next predictions. The other procedure was to use the same model, but to adjust the new residuals by the amount of the first residual after the maintenance action. The former procedure makes the linear model to continue more or less along the same path, and the slope constant coupled to time changes only slowly. The second procedure leads to a step change in the model data that reduces the slope coefficient. This latter procedure would, taken to its extremes, lead to a zero slope as track deterioration is followed by maintenance actions as long as the process is in operation.

For this reason, the former procedure is arguably better to use and the one recommended here. However, the degradation of the track may not be constant; it may de-accelerate as the degradation progresses, or it may accelerate. The slope and thus the predicted deterioration rate of the former method where old observations were compensated by a step change is too high, leading to increasingly large residuals. If this is generally true, or a particular artefact for the studied section needs further research, but if the slope change of the residuals is a general behaviour, both the reasons for the change of deterioration rate and the methods for more rapid calibration of the slope of the condition after maintenance needs further research.

5 Conclusions

One conclusion of this study is that out of the studied methods, the transformed twist variation was best modelled by a simple linear regression. The linear regression allowed for extrapolations of the behaviour and for monitoring the process.

The repeated measurements also allow for estimation of the measurement variation. A 95 % confidence interval of the measurement variation amounts to almost half of the totally measured variation. In such a case, the model predictions based on the location and evolvement of the property should be a much more reliable and useful source for maintenance action decisions.

6 Future Research

The different deterioration speeds before and after the maintenance actions should be of interest to study further, e.g. to support predictions of the track condition at an increased number of tamping versus a track renewal. A further development of the method could be that the prediction model should only be based on a selected number of the latest measurements, or that the slope of a linear model be based on a selected number of the latest measurements is a natural expansion of the proposed procedure.

If a particular deterioration process of a particular object changes, it may be important that the control chart method takes this change into account and continues to monitor deviations from the degradation path that the process is currently exhibiting. This means that the model must constantly be updated with new information about the current process dynamics. There is, of course a balance of how flexible the method should be allowed to be; too flexible and most deviations that should trigger an alarm would only be compensated for by that the model changes to follow in the direction towards the new observation.

The change of the degradation path may, on the other hand, be that what is interesting to monitor, rather than checking whether the last observation conveys to the current expectation or not. When the track condition is improved by maintenance actions, the degradation behaviour may be more or less rapid as when the improved object's properties were at the same level the last time. If, for instance, the track has been aligned and the ballast has been tamped, the deterioration rate of the track geometry may be much larger than the deterioration rate of a newly laid track with similar track alignment data, due to e.g. less benign ballast properties and so on. In fact, this deterioration effect of the ballast may limit the number of tamping actions before track renewal is necessary to achieve the desired track quality in a cost effective way. There may also be effects of initial settlement of the track bed so that the deterioration process for a new installation or a track renewal is higher than that of the tamped bed.

Acknowledgments We would like to thank Trafikverket (Swedish Transport Administration) and Luleå Railway Research Centre (JVTC) for supporting this work, both intellectually and financially. We are also grateful for insightful comments made by Arne Nissen that strengthened the study.

References

1. MacCarthy BL, Wasusri T (2002) A review of non-standard applications of statistical process control (SPC) charts. Int J Qual Reliab Manage 19:295–320
2. Montgomery DC (2008) Introduction to statistical process control. Wiley, New York
3. Bergquist B, Söderholm P (2014) Control charts supporting condition-based maintenance of linear railway infrastructure assets. In: Proceedings of the 3rd international workshop and congress on eMaintenance, Luleå, Sweden, 17–18 June 2014, pp 101–107
4. Bergquist B, Söderholm P (2014). Data analysis for condition-based railway infrastructure maintenance. E-pub ahead of print. Published online March 4, 2014. Quality and Reliability Engineering International
5. Van Beers WC, Kleijnen JP (2004). Kriging interpolation in simulation: a survey. In: Proceedings of the 2004 simulation conference, vol 1, Winter. IEEE

6. Akima H (1970) A new method of interpolation and smooth curve fitting based on local procedures. J ACM (JACM) 17(4):589–602
7. Banverket (1997) BVF 587.02—Spårlägeskontroll och kvalitetsnormer—Central mätvagn STRIX. Banverket, Borlänge
8. Banverket (1997) Z-647 97-11-15—Beskrivning av spårlägessystemet i mätvagn STRIX. Banverket, Borlänge
9. Bentley (2012) Bentley optram. http://www.bentley.com/en-US/Products/Bentley+Optram/. Accessed 20 Aug 2012
10. Trafikverket (2012) Optram. http://www.trafikverket.se/foretag/bygga-och-underhalla/jarnvag/system-verktyg-och-tjanster-for-jarnvagsjobb/optram/. Accessed 20 Aug 2012

Facilitating the Maintenance of Safety Cases

Omar Jaradat, Iain Bate and Sasikumar Punnekkat

Abstract Developers of some safety critical systems construct a safety case comprising both *safety evidence*, and a *safety argument* explaining that evidence. Safety cases are costly to produce, maintain and manage. Modularity has been introduced as a key to enable the reusability within safety cases and thus reduces their costs. The Industrial Avionics Working Group (IAWG) has proposed Modular Safety Cases as a means of containing the cost of change by dividing the safety case into a set of argument modules. IAWG's Modular Software Safety Case (MSSC) process facilitates handling system changes as a series of relatively small increments rather than occasional major updates. However, the process doesn't provide detailed guidelines or a clear example of how to handle the impact of these changes in the safety case. In this paper, we apply the main steps of MSSC process to a real safety critical system from industry. We show how the process can be aligned to ISO 26262 obligations for decomposing safety requirements. As part of this, we propose extensions to MSSC process for identifying the potential consequences of a system change (i.e., impact analysis), thus facilitating the maintenance of a safety case.

Keywords Safety case · Safety argument · Maintenance · Impact analysis · Change · IAWG MSSC

O. Jaradat (✉) · S. Punnekkat
Mälardalen University, Högskoleplan 1, 721 23 Västerås, Sweden
e-mail: omar.jaradat@mdh.se

S. Punnekkat
e-mail: sasikumar.punnekkat@mdh.se

I. Bate
University of York, Deramore Lane, York YO10 5GH, UK
e-mail: iain.bate@cs.york.ac.uk

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_25

1 Introduction

Constructing safety cases receives significant industrial attention as it is required for the certification process of many safety critical system domains. A safety case comprises both safety evidence (e.g. safety analyses, software inspections, or functional tests) and a safety argument explaining that evidence. Safety arguments show how system developers use each item of evidence to support claims, and how those claims, in turn, support broader claims about system behaviour, hazards addressed, and, ultimately, acceptable safety [1]. The production, management and evaluation of safety cases are considered difficult to achieve and time consuming. As an anecdotal example, the size of the preliminary safety case for surveillance on airport surfaces with ADS-B [2] is about 200 pages, and it is expected to grow as the operational safety case is created [3].

It is worth noting that a safety case is a living document that grows as the system grows. A safety case should be maintained as needed whenever some aspect of the system, its operation, its operating context, or its operational history changes.

Operational or environmental changes may invalidate a well-founded safety argument for different reasons as follows:

1. Changing the argument structure
2. Evidence is valid only in the operational and environmental context in which it is obtained, or to which it applies. During or after a system change, evidence might no longer support the developers' claims because it could reflect old development artefacts or old assumptions about operation or the operating environment
3. In the updated system, existing safety claims might be nonsense, no longer reflect operational intent, or they might be contradicted by new data

The certification process must be repeated after applying changes to an already certified system (i.e., re-certification). In other words, the safety case of the certified system should show that the system is acceptably safe to operate in its intended context after applying the changes. In order to achieve the re-certification, a safety argument should be maintained by determining whether the item of evidence still supports the claims made about it, check whether new or updated safety requirements are reflected in the argument, and review the overall logic of the argument. The main problem though is that the elements of the safety argument (i.e., safety goals, evidence, argument and the operating context) are highly interdependent so that what can be seen as a minor change in the argument may have a major impact to the contents and the structure of that argument [4]. Hence, maintaining a safety argument requires high awareness of the dependencies among its contents and how a change to one part may invalidate other parts. Without this vital awareness, a developer performing impact analysis might not notice that a change has compromised system safety. The Ariane 5 rocket which crashed 40 seconds after take-off in 1996 is a costly example of omitting affected parts of a system due to a change. Ariane 5 inertial reference system (SRI) tried to stuff a 64-bit number into a

16-bit space which led to a conversion error. This part of the system was reused from an older version of the SRI that was implemented for Ariane 4 rocket. Seemingly, an assumption was made as since the code was successfully used in an older version of the system then it is suitable to be reused for the newer version [5]. Hence, system developers focused on more complex parts of the system and no attention was paid to the out-of-date code or to any related assumption.

A fundamental step prior to update a safety case due to a change is to assess the impact of this change in the safety argument. This is referred to as safety case impact analysis. It is probably clearer now how the continuous maintenance efforts to keep the safety case always up-to-date add more burden on top of the discussed difficulties above. Moreover, the cost of change has become a major part of the cost of ownership of a system [6].

As a response to these challenges, an ambition emerged to modularize safety cases by applying the principles of software architecture and design to the safety case domain. The main idea of the modularity is to align boundaries of safety case modules with design boundaries to contain changes. Having done that, a change to a design element should then affect the corresponding safety case module, and not impact the entire safety argument [6].

To this end, the Industrial Avionics Working Group (IAWG) represented by a team of highly experienced engineers, experts in software development and safety assurance, defined the Modular Software Safety Case (MSSC) process [7] as a means for containing the cost of change by dividing the safety case into a set of argument modules. The process has been refined through experience gained from large-scale trial applications of the prototype process, and further trials of the refined process. MSSC process establishes component traceability mechanism between system design elements and safety argument modules by using the concepts of Dependency-Guarantee Relationship (DGR) and Dependency-Guarantee Contract (DGC). The former is to highlight, and describe, safety-related properties and behaviour of a single design element. In other words, DGRs capture the relationships between input and output ports for each design element. A DGC, however, is used to match one design element's dependencies with another design element's guarantees [8].

The contributions of this paper are as follows: demonstrating how to apply the IAWG MSSC process. More specifically, apply the process to the Fuel Level Estimation System (FLES), which is a real safety critical system that was implemented by Scania AB—a major Swedish automotive industry manufacturer—to show (1) how the DGR and DGC concepts can be used to capture the safety requirements of the FLES, (2) how these two concepts can be used to build a safety case in conformance to the requisites of ISO 26262 for certification, and (3) extending IAWG's DGC to improve the impact analysis process thus facilitating the maintenance of safety cases.

This paper is composed of four further sections. In Sect. 2 we present background information. In Sect. 3 we present the IAWG MSSC process. In Sect. 4 we use the FLES to demonstrate the application of the IAWG MSSC process. Finally, in Sect. 5 we draw conclusions and identify future work.

2 Background

This section presents background information about the safety standard ISO 26262, the Goal Structuring Notation (GSN), safety case maintenance and current challenges, and an approach to maintaining safety case evidence after a system change.

2.1 The Safety Standard ISO 26262

The rationale behind the selection of this standard for this work is that it is functional safety standard was adapted for automotive electric/electronic systems that Scania is working to qualify for its certification stamp. Since FLES is one of other systems in Scania's trucks, it is very appropriate to consider ISO 26262 for the given example in this paper.

ISO 26262 regulates the automotive domain, more specifically, the standard is intended to be applied to safety-related systems that include one or more electrical and/or electronic systems and that are installed in series production passenger cars with a maximum gross vehicle mass up to 3500 kg [9]. In this subsection, however, we focus only on the part of the standard that regulates the decomposition of safety requirements. The following parts are summarized descriptions of the safety requirements decomposition directly from ISO 26262 guidelines:

1. Successively after identifying hazards, the standard recommends to formulate the Safety Goals (SGs) related to the prevention or mitigation of the hazardous events, in order to avoid unreasonable risk. Basically, hazard analysis, risk assessment and Automotive Safety Integrity Level (ASIL) are used to determine the safety goals such that an unreasonable risk is avoided. The standard defines a safety goal as a top-level safety requirement resultant of the hazard analysis and risk assessment. Safety goals are not expressed in terms of technological solutions, but in terms of functional objectives [9].
2. Identification of safety goals leads to the functional safety concept. The objective of the functional safety concept is to derive the Functional Safety Requirements, from the safety goals, and to allocate them to the preliminary architectural elements. To comply with the safety goals, the functional safety concept contains safety measures, including the safety mechanisms, to be implemented in the item's architectural elements and specified in the functional safety requirements. The standard defines a functional safety requirement as a specification of implementation-independent safety behaviour, or implementation-independent safety measure, including its safety-related attributes [9].
3. Finally, both the functional concept and the preliminary architectural assumptions lead to the technical safety concept. The first objective of this concept is to specify the Technical Safety Requirements and their allocation to system elements for implementation by the system design. The second objective is to verify through analysis that the technical safety requirements comply with the functional safety

requirements. The standard defines a technical safety requirement as a requirement derived for implementation of associated functional safety requirements [9].

2.2 The Goal Structuring Notation (GSN)

A safety argument organizes and communicates a safety case, showing how the items of safety evidence are related and collectively demonstrate that a system is acceptably safe to operate in a particular context. The GSN [10] provides a graphical means of communicating (1) safety argument elements, claims (goals), argument logic (strategies), assumptions, context, evidence (solutions), and (2) the relationships between these elements. The principal symbols of the notation are shown in Fig. 1 (with example instances of each concept).

A goal structure shows how goals are successively broken down into ("solved by") sub-goals until a point is reached where claims can be supported by direct reference to evidence. Using the GSN, it is also possible to clarify the argument strategies adopted (i.e., how the premises imply the conclusion), the rationale for the approach (assumptions, justifications) and the context in which goals are stated. It is worth noting that GSN has been extended to enable modularity in a safety case (i.e., module-based development of the safety case). Hence, modular GSN enables the partitioning of a safety case into an interconnected set of modules.

2.3 Safety Case Maintenance and Current Challenges

A safety case is a living document that should be maintained whenever some aspect of the system, its operation, its operating context, or its operational history changes. In this paper, the process of updating the safety case after implementing a system change is referred to as safety case maintenance.

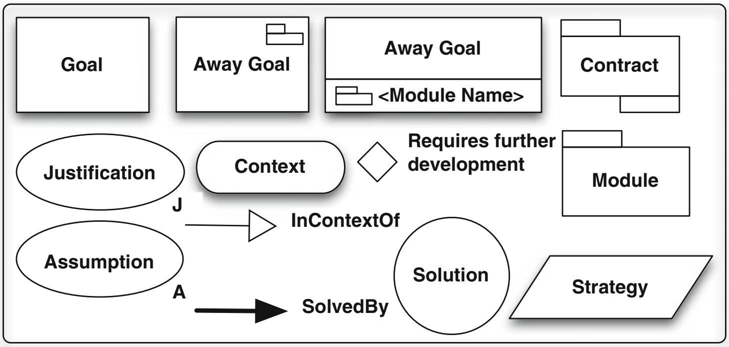


Fig. 1 Overview of Goal Structuring Notation (GSN)

Developers of safety critical systems experience difficulties in safety case maintenance after implementing a system change. One of the main difficulties is identifying the impacted parts in the safety argument. The traceability between a system design and the corresponding safety argument contents, and the dependency among the contents of safety argument are considered two main burdens that encounter the identification of the impacted parts in an argument. Moreover, individual systems tend to become more complex as they are designed and constructed, this increasing complexity, as well as, the number of evidence items in a safety argument can exacerbate the maintenance difficulties. Any approach intends to manage safety argument due to system changes should consider:

1. A means for clearly capturing the underlying rationale of the safety argument in order to assess the impact of change on all parts of the argument
2. A traceability mechanism between a system domain and the safety argument to support the ability to track the changed part from the system design down to the corresponding affected part in the safety argument
3. Mechanisms to structure the argument so as to contain the impact of changes

The use of the GSN approach helps to produce well-structured arguments that clearly demonstrate the argument elements and their interdependencies (the relationships between the argument claims and evidence) [4, 11, 12]. Using GSN makes capturing the underlying rationale of the argument easier, which will in turn, help to scope areas affected by a particular change and thus helps the developers to mechanically propagate the change through the goal structure. However, GSN does not tell if the suspect elements of the argument in question are still valid. For example, having made a change to a model we must ask whether goals articulated over that model are still valid. Expert judgment, therefore, is still required in order to answer such questions. Hence, using GSN does not directly help to maintain the argument after a change, but it can more easily determine the questions to be asked to do so [12].

Current standards and analysis techniques assume a top-down development approach to system design. For component-based systems, monolithic evidence produced via these approaches is difficult to maintain those systems because it is hard to match a safety argument that has a different structure than the system design structure. However, safety is a system level property and assuring this property requires every piece of evidence generated for each component to be linked and compared to demonstrate consistency [7]. One may think that the matching (i.e., optimal level of traceability) can be achieved by designing a safety argument structure to be similar to the system design structure, where a clear one-to-one mapping of a system design component to a safety argument module can be established (see Fig. 2).

Theoretically, a one-to-one mapping may facilitate tracking down the components of a system design to the safety argument, but it is impractical due to four key factors: (1) modularity of evidence, (2) modularity of the system, (3) process demarcation (e.g., ISO 26262 items [9]), and (4) organisational structure (e.g., who is working on what). These factors have a significant influence when deciding upon the safety argument structure.

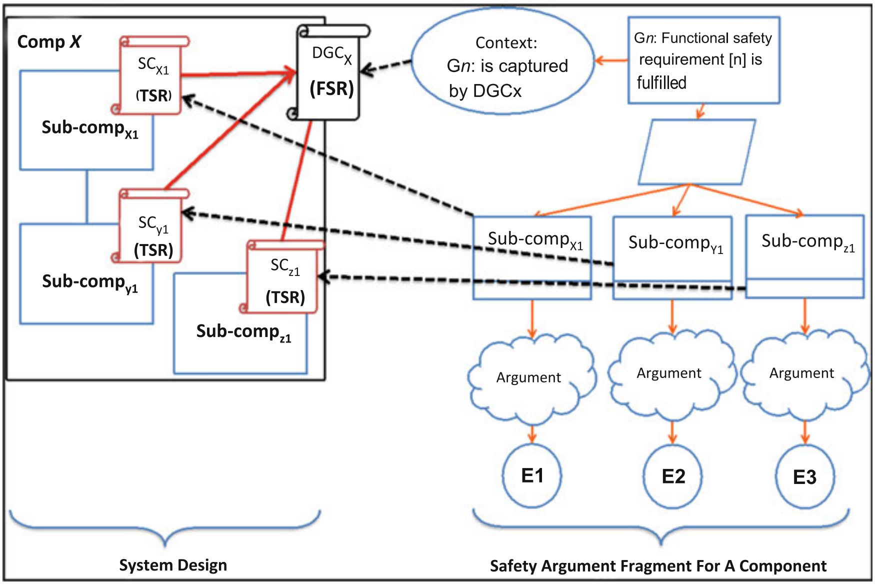


Fig. 2 An illustration of the relationship between a system design and its safety argument

Enabling component and evidence traceability is very useful to analyse the impact of change on a safety argument, and eventually, facilitates the overall maintenance of the safety case. This paper deals with two forms of traceability: component (i.e. safety argument fragment to system design component) and evidence (i.e. safety argument fragment to supporting evidence). However, to the best of our knowledge there are no supporting process or method that provides detailed steps of how to analyse the impact of a change on a safety case using component or evidence traceability. That said there are well-regarded industry-lead initiatives that assume such methods exist. MSSC Process is one such example.

In this paper, we use the word "traceability" to indicate two different things. Firstly, we refer to the ability to relate safety argument fragments to system design components as component traceability mechanism (through a safety argument). Secondly, we refer to the ability to relate safety argument evidence across system's artefacts as evidence traceability.

2.4 *Maintaining Safety Case Evidence After a System Change*

In our previous work [1], we proposed a new approach to facilitating safety case change impact analysis. In the approach, automated analysis of information given as annotations to a safety argument (recorded in the GSN) highlight suspect safety evidence to bring it to engineer's attention. We proposed annotating each reference to a development artefact (e.g. an architecture specification) in a goal or context element with an artefact version number.

We also proposed annotating each solution element with:

1. An evidence version number
2. An input manifest identifying the inputs (including version) from which the evidence was produced
3. The lifecycle phase during which the evidence obtained (e.g. Software Architecture Design)
4. A safety standard reference to the clause in the applicable standard (if any) requiring the evidence (and setting out safety integrity level requirements)

With this data, we can perform a number of automated checks to identify items of evidence impacted by a change. For example:

1. We can determine when two different versions of the same item of evidence are cited in the same argument
2. We can identify out-of-date evidence by searching for input manifests $m = \{(a1, v1),\ldots, (an, vn)\}$ and artefact versions (a, v) such that $\exists i \bullet a = ai \wedge v > vi$
3. Where we know a particular artefact has changed, we can search for input manifests containing old versions

If we had further information which inputs were used to produce each input listed in each input manifest, each input that was used to produce those, and so on, we could extend checks (2) and (3) above to indirect inputs. For example, suppose that life testing is used to establish the reliability of a component, that this component and its reliability appear in a Failure Modes and Effects Analysis (FMEA), and that the FMEA results are used in a Fault Tree Analysis (FTA). With the additional information, we could compute a closure of the FTA's input manifest that would include the life testing results. Other analyses may be possible. For example, we suggest storing the safety standard reference to facilitate analysis of impacts that change the safety integrity level of a requirement.

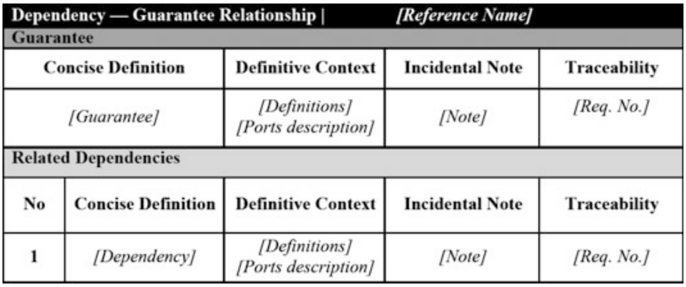
3 Modular Software Safety Case (MSSC) Process

IAWG has proposed Modular Safety Cases as a means of containing the cost of change by dividing the safety case into a set of argument modules. IAWG's MSSC process facilitates handling system changes as a series of relatively small increments rather than occasional major updates (i.e., incremental certification). MSSC process manages system changes by breaking down a system into blocks. The process defines the block as an identifiable part (or group of parts) of the Software implementation that is chosen by the safety case architect to be the subject of a safety case module. Blocks cover all parts of a system design where each block may correspond to a single or multiple software component or unit of design, but it is subject to only one dedicated safety case module. In other words, each system block has one-to-one relationship with a safety argument module [7].

The process establishes component traceability mechanism between system blocks and safety argument modules by using the concepts of DGR and DGC as shown in Figs. 3 and 4, respectively. The former is to highlight and describe safety-related properties and behaviour of a system block. In other words, a DGR captures the relationships between input and output ports for each design block. A DGC, however, is used to match one block's dependencies with another block's guarantees [7, 13]. Creating DGCs leads to the creation of a 'daisy chain' as a dependency in one block and a guarantee offered by another, whose associated dependencies are supported by further guarantees, and so on [13].

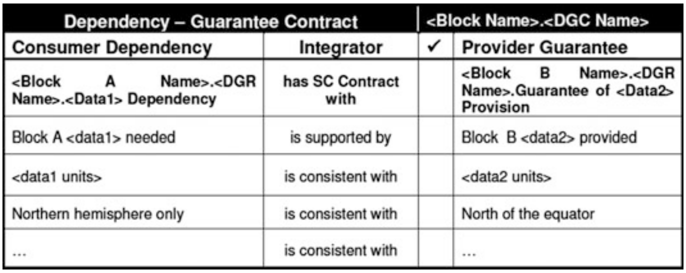
MSSC process is very dependent on the anticipated changes that should be identified in the first step of the process. The anticipated change scenarios will bring the highly likely changeable parts in the system to developer's attention.

These scenarios are considered by system developers so that they can manage the containment of the impact of these changes in the system blocks boundaries more efficiently. Having done this, the impact of a change in one safety argument module will hopefully not propagate into another module, but it might impose one (or more) safety case contract update, and even if it is then the cost of changes can be minimised.



<table>
<tr><th colspan="3">Dependency — Guarantee Relationship |</th><th colspan="2"><em>[Reference Name]</em></th></tr>
<tr><td colspan="5"><strong>Guarantee</strong></td></tr>
<tr><td colspan="2"><strong>Concise Definition</strong></td><td><strong>Definitive Context</strong></td><td><strong>Incidental Note</strong></td><td><strong>Traceability</strong></td></tr>
<tr><td colspan="2"><em>[Guarantee]</em></td><td><em>[Definitions]</em>
<em>[Ports description]</em></td><td><em>[Note]</em></td><td><em>[Req. No.]</em></td></tr>
<tr><td colspan="5"><strong>Related Dependencies</strong></td></tr>
<tr><td><strong>No</strong></td><td><strong>Concise Definition</strong></td><td><strong>Definitive Context</strong></td><td><strong>Incidental Note</strong></td><td><strong>Traceability</strong></td></tr>
<tr><td>1</td><td><em>[Dependency]</em></td><td><em>[Definitions]</em>
<em>[Ports description]</em></td><td><em>[Note]</em></td><td><em>[Req. No.]</em></td></tr>
</table>

Fig. 3 A DGR tabular representation



Dependency – Guarantee Contract			<Block Name>.<DGC Name>
Consumer Dependency	**Integrator**	✓	**Provider Guarantee**
<Block A Name>.<DGR Name>.<Data1> Dependency	has SC Contract with		<Block B Name>.<DGR Name>.Guarantee of <Data2> Provision
Block A <data1> needed	is supported by		Block B <data2> provided
<data1 units>	is consistent with		<data2 units>
Northern hemisphere only	is consistent with		North of the equator
...	is consistent with		...

Fig. 4 A DGC tabular representation

It is very important to distinguish between a DGC and a safety case contract. The former captures the required link between a dependency declared in one DGR and a satisfying guarantee provided by another. Hence, DGCs are created on the system design level. A safety case contract, however, is used to describe the linkage between a consumer goal in one Safety Case Module and a provider goal in another [7]. This is formed through the new GSN extension for modularity.

Figure 5 shows an example to describe the relationships between system blocks, DGR, DGC, safety case contract and the safety case architecture. It is worth noting that DGCs may be linked to safety case contracts.

The following is a list summarises MSSC process's steps [7]:

Step 1 **Analyse the product lifecycle**: it is important to predict the potential change scenarios over the projected system lifetime. One reason for that is because change scenarios will help assess the potential benefits that may be achieved through modular certification. If as a result of the analysis there are no changes expected, then the full benefits of modular certification may not be realised, and it may therefore be decided not to adopt a modular approach [13].

Step 2 **Optimise software design and safety case architecture**: since each system block is subject to safety case module. First, we need to divide the system into blocks and form public interfaces for the block safety case modules. All elements of the system are split into blocks and each corresponding block safety case module should present an argument about the safety-related behaviour of that block. Second, other necessary modules will be added, for example, software safety requirements, software system

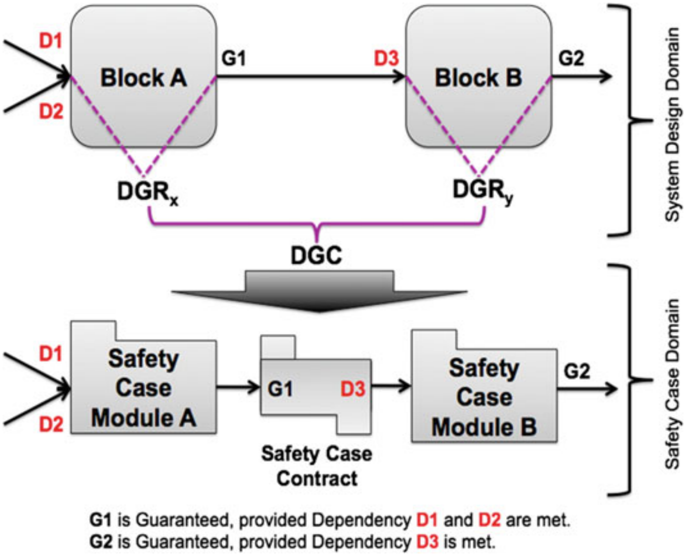


Fig. 5 Linking blocks using DGRs and DGCs

wide issues module, configuration data module, safety case contract modules, etc. Finally, we should define safety case integration modules—these provide the argument about the combined behaviour of interdependent safety case modules [7].

Step 3 **Construct safety case modules**: A hazard mitigation argument should be formed and derived safety requirements are directed to SW blocks safety case modules. The guaranteed behaviour offered by each block in support of these is captured, along with dependencies on other blocks. A Block Safety Case Module is constructed providing argument and evidence for each Block based on the Guarantees and Dependencies [7].

Step 4 **Integrate safety case modules**: the safety case modules are integrated so that claims requiring support in one Safety Case Module are linked to claims providing that support in others. This step of the process results in a fully integrated Safety Case [7].

Step 5 **Assess/Improve change impact**: when a system change is implemented, the impact on the design modules and associated Safety Case Modules is assessed [7].

Step 6 **Reconstruct safety case modules**

Step 7 **Reintegrate safety case modules**

Step 8 **Appraise the safety case**

The guidance of MSSC process [7] does not show detailed information about how to follow some steps including the impact analysis part. The provided example by the process abstracts the impact analysis step and shows its results only. The main work in this paper is not to consider all parts of MSSC process to give a full example on how to apply them but we rather focus on the impact analysis part and necessary prerequisite steps only.

4 Illustrative Example: Fuel Level Estimation System (FLES)

In our previous work [14, 15], we used FLES as a specimen system to illustrate the contribution of the architectural model checking to conduct preliminary safety assessment in line with the safety standard ISO 26262.

We used the Architecture Analysis & Design Language (AADL) to model the system as shown in Fig. 6. In our current work we reuse the description as well as the AADL of FLES to partially apply MSSC process. We also propose a system change scenario and examine how the method helps to highlight the affected safety argument elements.

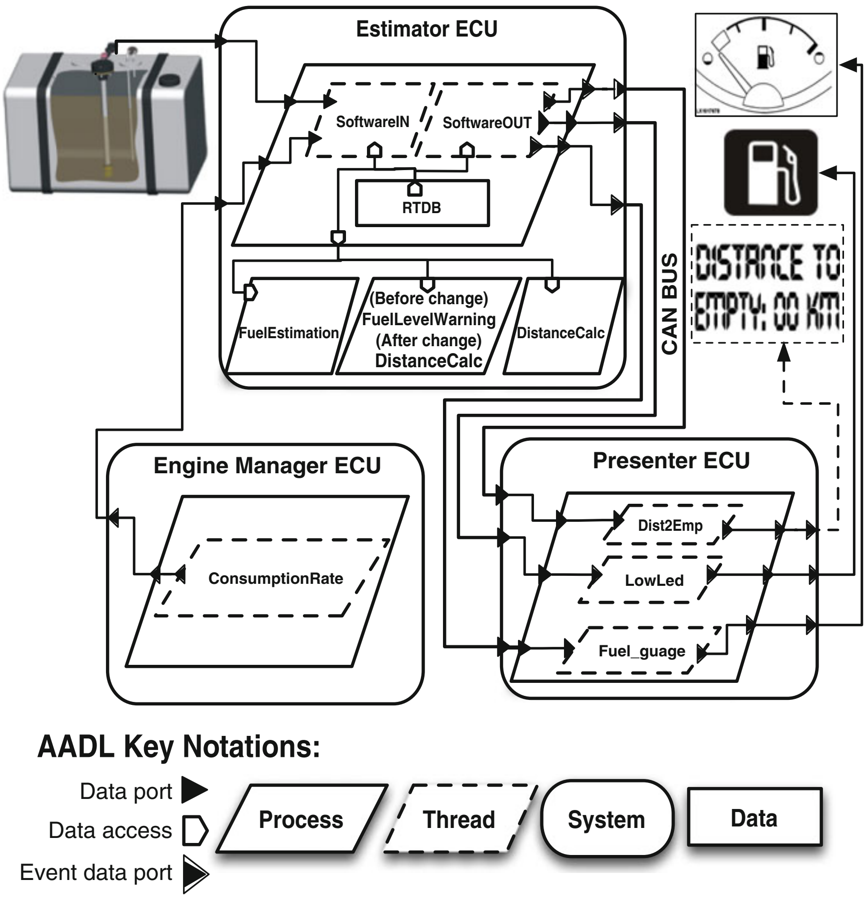


Fig. 6 An AADL representation of Estimator's software architecture

4.1 FLES Description

4.1.1 FLES Technical Details

FLES estimates the volume of fuel in a heavy road vehicle's tank and presents this information to the driver through a dashboard mounted fuel gauge. Additionally, the system must warn the driver when this volume falls below a predefined threshold. This system is considered safety critical because its failure could lead to loss of control of the vehicle. For example, if there is less fuel remaining than the driver thinks, the vehicle might run out, bringing it to an unexpected halt, which can be hazardous in certain contexts. As well as bringing the vehicle to a halt, the power steering and braking mechanisms could also fail. These failures would compromise vehicle controllability and could also lead to a crash.

Fuel volume is estimated using a float sensor in the fuel tank. As the position of the float is affected by vehicle motion (negotiating steep hills, sharp bends, or rough terrain), the system has some challenging issues to be tackled within its design. The system must process this signal so that at all times the gauge displays an accurate measurement of the total volume of fuel remaining. The sensed value is sent to the *Estimator ECU*. An Analogue to Digital Converter (ADC) is used to convert and then the *SoftwareIN* thread reads the sensed fuel float position from the ADC and stores it in the real-time database *RTDB*. *FuelEstimation* reads this sensor value and computes an estimate of the current fuel volume in litres. When the vehicle might be moving (i.e., its parking brake is not set), the *FuelEstimation* thread uses a Kalman filter algorithm to reduce the noise introduced by vehicle motion. This algorithm requires the recent history of fuel volume estimates to be stored. *FuelEstimation* outputs a smoothed fuel volume estimate to the *RTDB*. *FuelLevelWarning* then reads this estimate, compares it to the low-fuel warning threshold (i.e., < 7 % of the fuel tank capacity), and writes the low-fuel warning status to the *RTDB*. *SoftwareOUT* reads the fuel volume and low-fuel warning status from the *RTDB* and sends these over the Controller Area Network (CAN) bus to the *Presenter ECU*. The *Presenter ECU* adjusts the actuators (i.e., fuel gauge and low-fuel lamp) on the dashboard according to the received values.

4.1.2 FLES Safety Analysis

Hazard analysis and risk assessment made for FLES led to one hazard identification: "*Unannunciated lack of fuel*". Unannunciated is interpreted as (1) fuel estimates and low-fuel warning are not displayed at all, and (2) it is displayed incorrectly since the estimates are not identical to the real amount of fuel in the vehicle's tank. The determined ASIL for the fuel level estimation system is "C".

The derived safety requirements to mitigate the hazard are decomposed as recommended by ISO 26262 as follows:

1. ***Safety goals***: two safety goals were derived
   a. *SG1.0ImplAssur*: Vehicle's driver shall be constantly aware of the actual remaining fuel in the tank whenever the engine is in operation
   b. *SG2.0ImplAssur*: Vehicle's driver shall be warned when the fuel level is low and the engine is in operation
2. ***Functional Safety Requirements (FSR)***:
   Two functional safety requirements were identified to satisfy *SG1.0ImplAssur*:
   a. *ConFSR1.0.1.0*: A fuel gauge should promptly annunciate the actual fuel amount in the tank whenever the engine is in operation
   b. *ConFSR1.0.2.0*: The fuel gauge shall not display a fuel estimate that deviates more that 5 % from the actual fuel volume in the tank
   One functional safety requirement was identified to satisfy *SG2.0ImplAssur*:

Table 1 A Subset of the identified TSRs for FLES

FSR ID	TSR ID	Description
FSR1.0.1.0	*F1010TSR1*	The *FuelEstimation* thread shall provide the *totalFuelLevel* value
FSR1.0.1.0	*F1010TSR2*	The *SoftwareOUT* shall send the *totalFuelLevel* value to the *Presenter*
FSR2.0.1.0	*F2010TSR1*	The *FuelLevelWarning* thread shall provide *lowFuelWarning* value
FSR2.0.1.0	*F2010TSR2*	The *SoftwareOUT* shall send the *lowFuelWarning* value to the *Presenter*

c. *ConFSR2.0.1.0*: A fuel-low warning lamp should be promptly turned ON when the fuel level in the tank falls below a certain level whenever the engine is in operation

3. ***Technical Safety Requirements (TSR)***: There is a large set of technical safety requirements that was identified to specify the functional safety requirements. The work of the paper, however, considers the minimum set of technical safety requirements that specify *ConFSR1.0.1.0* and *ConFSR2.0.1.0* as shown in Table 1.

4.2 Applying the IAWG MSSC Process

A list of anticipated change scenarios during FLES's lifetime is required. This list may help assessing the potential benefits that may be achieved through modular certification. In this section, we present the details of the various MSSC process steps with respect to FLES:

4.2.1 Analyse the Product Lifecycle and Identify Change Scenarios

We assume one potential change for FLES. The *Distance To Empty* feature might be added to FLES. The role of this anticipated change is to determine the distance (Km) that a vehicle can drive before it runs out of fuel. This new feature is dependent on (1) the estimation of the current fuel amount in the tank (L), and (2) the fuel consumption rate (L/Km) in the engine. Technically, this intended feature will be added as a new thread in the *Estimator ECU*. This thread should read the output of the *FuelEstimation* thread, as well as, the output of the *ConsumptionRate* thread that is implemented in the *EngineManager ECU*. To avoid dealing with timing and memory budgets, FLES engineers expect to remove the *FuelLevelWarning* thread and move the task it contains to the *FuelEstimation* thread (i.e., merge the two threads into one). Since the safety margin of the

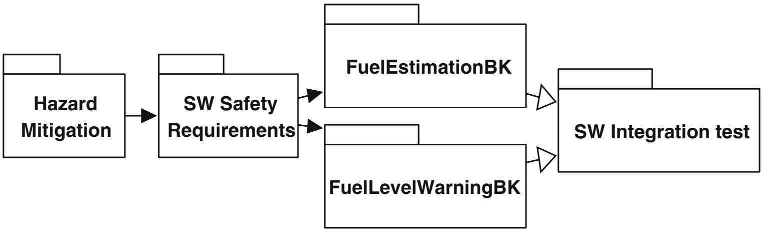


Fig. 7 FLES safety case architecture

FuelEstimation thread allows adding a new task, the timing and memory budget for the thread will remain the same even after the merge. On the other hand, the new *DistanceCalc* thread will take the timing and memory budget, and the priority of the removed *FuelLevelWarning* thread. The same arrangements will be applied to the threads in the *Presenter ECU*.

4.2.2 Optimise Software Design and Safety Case Architecture (Define the Safety Case Architecture)

For the sake of simplicity, we do not define a full set of the safety case modules, but we rather define the basic modules that are sufficient to make the example. We focus on the *Estimator* in our example by dividing it into two software blocks, namely, *FuelEstimationBK* and *FuelLevelWarningBk*. Each of them represents a safety case module. Additionally, we construct *Hazard Mitigation, SW Safety Requirements and SW Integration test* modules (as shown in Fig. 7).

4.2.3 Construct Safety Case Modules, and Integrate Safety Case Modules

We merge these two steps for the sake of simplicity. We identify the required DGRs of the *FuelEstimationBK* and *FuelLevelWarningBk* blocks. We also construct the *Hazard Mitigation, SW Safety requirements*, *FuelEstimationBK*, *FuelLevelWarningBk*, and *Software Integration test* safety case modules.

Table 2 shows one DGR for the software block FuelEstimationBK in which the block (i.e., represented as thread) guarantees that it can provide the estimated fuel level volume in the tank *totalFuelLevel* if the three dependencies are met. Table 3 shows one DGR for the software block *FuelLevelWarningBK* in which the block (i.e., represented as thread) guarantees that it can tell if the fuel is low or not (*lowFuelLevelWarning* is *True* if the fuel is below 7 % of the tank capacity and *False* if the fuel is not) once the four related dependencies are met.

In Fig. 8, we construct the hazard mitigation argument. Basically, *MitigationHazard1* goal is supported by implementing and assuring the two safety

Table 2 DGR FuelEstimationBK

Dependencies—guarantee relationship | FuelEstimation*BK.G5*

Guarantee

Concise definition	Definitive context	Incidental note	Traceability
Provides the *totalFuelLevel* value	The *totalFuelLevel* value is sent on port *SetSensorValue*		***F1010TSR1***
	The totalFuelLevel format is defined by FLES {Interface Specification}		

Related dependencies

N	Concise definition	Definitive context	Incidental note	Traceability
1	*FuelLevelSensor* is received via port *GetSetSensorValue*	*FuelLevelSensor* format is defined by FLES {Interface Specification}		*F3010TSR8*
2	*SetSensorValue* port is available	The port behaviour is as defined in the FLES {Interface Description}		*F3010TSR9*
3	*FuelEstimation* is correctly configured	Is executing and has completed configuration		*F4010TSR5*

Table 3 DGR FuelLevelWarningBK

Dependencies—guarantee relationship | FuelLevelWarningBK.G1

Guarantee

Concise definition	Definitive context	Incidental note	Traceability
Provides the *lowFuelLevelWarning* value	The *lowFuelLevelWarning* value is sent on port *setlowFuelLevelWarning*		*F2010TSR1*
	lowFuelLevelWarning format is defined by FLES {Interface Specification}		

Related dependencies

N	Concise definition	Definitive context	Incidental note	Traceability
1	*totalFuelLevel* is received via port *GetEstimatedFuelLevelValue_2*	*FuelLevelSensor* format is defined by FLES {Interface Specification}		*F1010TSR1*
2	*set*low*FuelLevelWarning* port is available	The port behaviour is as defined in the FLES {Interface Description}		*F3010TSR9*
3	*GetEstimatedFuelLevelValue_2* port is available			*F4010TSR5*
4	*FuelEstimation* is correctly configured	Is executing and has completed configuration		*F4010TSR7*

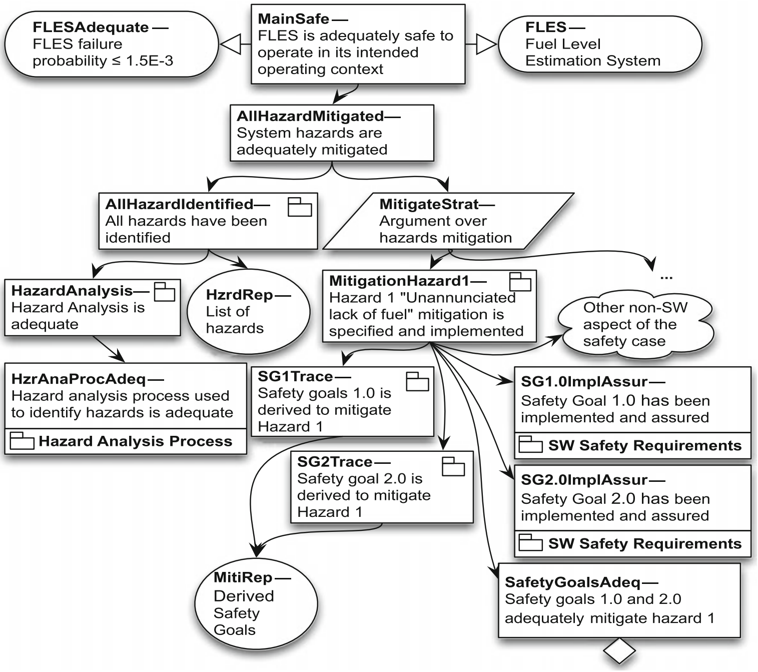

FLESAdequate —
FLES failure probability ≤ 1.5E-3
MainSafe —
FLES is adequately safe to operate in its intended operating context
FLES —
Fuel Level Estimation System
AllHazardMitigated—
System hazards are adequately mitigated
AllHazardIdentified—
All hazards have been identified
MitigateStrat—
Argument over hazards mitigation
...
HazardAnalysis—
Hazard Analysis is adequate
HzrdRep—
List of hazards
MitigationHazard1—
Hazard 1 "Unannunciated lack of fuel" mitigation is specified and implemented
Other non-SW aspect of the safety case
HzrAnaProcAdeq—
Hazard analysis process used to identify hazards is adequate
Hazard Analysis Process
SG1Trace—
Safety goals 1.0 is derived to mitigate Hazard 1
SG1.0ImplAssur—
Safety Goal 1.0 has been implemented and assured
SW Safety Requirements
SG2Trace—
Safety goal 2.0 is derived to mitigate Hazard 1
SG2.0ImplAssur—
Safety Goal 2.0 has been implemented and assured
SW Safety Requirements
MitiRep—
Derived Safety Goals
SafetyGoalsAdeq—
Safety goals 1.0 and 2.0 adequately mitigate hazard 1

Fig. 8 Hazard mitigation safety case module of FLES

goals that were derived to mitigate it. The safety goals are represented by the two separated away goals *SG1.0ImplAssur,* and *SG2.0ImplAssur*. These goals also represent the integration between *Hazard Mitigation* safety case module and *SW Safety Requirements* (see Fig. 9).

In *FuelLevelWarning.BK* Safety case module (see Fig. 10), we show how arguing over the dependencies supports the guarantee that is represented by *FuelLevelWarningBK.G1*. The argument module uses *FuelEstimationBK.G5* as a dependency to support the guarantee. *FuelEstimationBK.G5* also relies on a set of dependencies to be guaranteed. Figure 11 shows an argument fragment of the *SW Integration test* safety case module. The objective of the module is to argue over the integration of the software elements within the *Estimator ECU*.

The *FuelLevelWarningBK.G1* DGR shows that in order for *FuelLevelWarningBK* being able to fulfil the TSR *F2010TSR1* it requires the TSR *F1010TSR1,* which is guaranteed by a different DGR (i.e., *FuelEstimationBK.G5)*. Here lies the importance of the DGC as it matches such dependencies. Table 4 shows a DGC that matches *F2010TSR1* to *F1010TSR1*. MSSC process requires performing the integration of safety case modules by using a safety case contract module. The latter uses a DGC to set out the matching between the DGRs of the goals involved. However, since our work is more focused on facilitating the impact analysis within the blocks, we do not use safety case contracts in this example thus

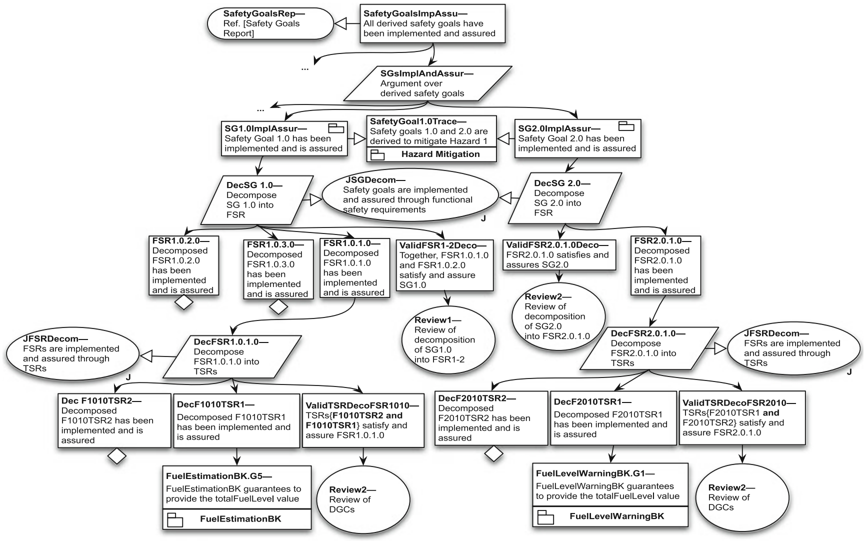

SafetyGoalsRep—
Ref. [Safety Goals Report]
SafetyGoalsImpAssu—
All derived safety goals have been implemented and assured
...
SGsImplAndAssur—
Argument over derived safety goals
...
SG1.0ImplAssur—
Safety Goal 1.0 has been implemented and is assured
SafetyGoal1.0Trace—
Safety goals 1.0 and 2.0 are derived to mitigate Hazard 1
Hazard Mitigation
SG2.0ImplAssur—
Safety Goal 2.0 has been implemented and is assured
DecSG 1.0—
Decompose SG 1.0 into FSR
JSGDecom—
Safety goals are implemented and assured through functional safety requirements
J
DecSG 2.0—
Decompose SG 2.0 into FSR
FSR1.0.2.0—
Decomposed FSR1.0.2.0 has been implemented and is assured
FSR1.0.3.0—
Decomposed FSR1.0.3.0 has been implemented and is assured
FSR1.0.1.0—
Decomposed FSR1.0.1.0 has been implemented and is assured
ValidFSR1-2Deco—
Together, FSR1.0.1.0 and FSR1.0.2.0 satisfy and assure SG1.0
ValidFSR2.0.1.0Deco—
FSR2.0.1.0 satisfies and assures SG2.0
FSR2.0.1.0—
Decomposed FSR2.0.1.0 has been implemented and is assured
Review1—
Review of decomposition of SG1.0 into FSR1-2
Review2—
Review of decomposition of SG2.0 into FSR2.0.1.0
JFSRDecom—
FSRs are implemented and assured through TSRs
J
DecFSR1.0.1.0—
Decompose FSR1.0.1.0 into TSRs
DecFSR2.0.1.0—
Decompose FSR2.0.1.0 into TSRs
JFSRDecom—
FSRs are implemented and assured through TSRs
J
Dec F1010TSR2—
Decomposed F1010TSR2 has been implemented and is assured
DecF1010TSR1—
Decomposed F1010TSR1 has been implemented and is assured
ValidTSRDecoFSR1010—
TSRs{F1010TSR2 and F1010TSR1} satisfy and assure FSR1.0.1.0
DecF2010TSR2—
Decomposed F2010TSR2 has been implemented and is assured
DecF2010TSR1—
Decomposed F2010TSR1 has been implemented and is assured
ValidTSRDecoFSR2010—
TSRs{F2010TSR1 and F2010TSR2} satisfy and assure FSR2.0.1.0
FuelEstimationBK.G5—
FuelEstimationBK guarantees to provide the totalFuelLevel value
FuelEstimationBK
Review2—
Review of DGCs
FuelLevelWarningBK.G1—
FuelLevelWarningBK guarantees to provide the totalFuelLevel value
FuelLevelWarningBK
Review2—
Review of DGCs

Fig. 9 SW Safety Requirements safety case module

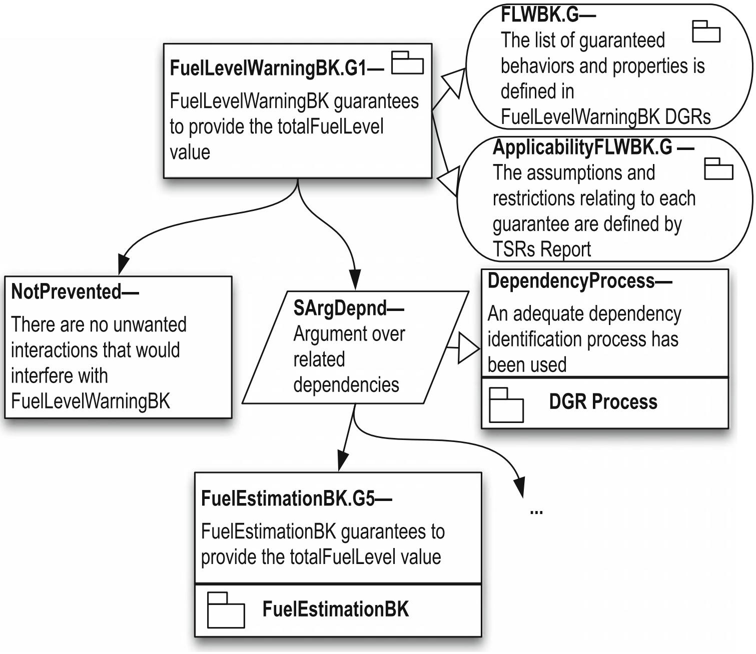

FLWBK.G—
The list of guaranteed behaviors and properties is defined in FuelLevelWarningBK DGRs
FuelLevelWarningBK.G1—
FuelLevelWarningBK guarantees to provide the totalFuelLevel value
ApplicabilityFLWBK.G —
The assumptions and restrictions relating to each guarantee are defined by TSRs Report
NotPrevented—
There are no unwanted interactions that would interfere with FuelLevelWarningBK
SArgDepnd—
Argument over related dependencies
DependencyProcess—
An adequate dependency identification process has been used
DGR Process
FuelEstimationBK.G5—
FuelEstimationBK guarantees to provide the totalFuelLevel value
FuelEstimationBK
...

Fig. 10 An argument fragment of FuelLevelWarning.BK safety case module

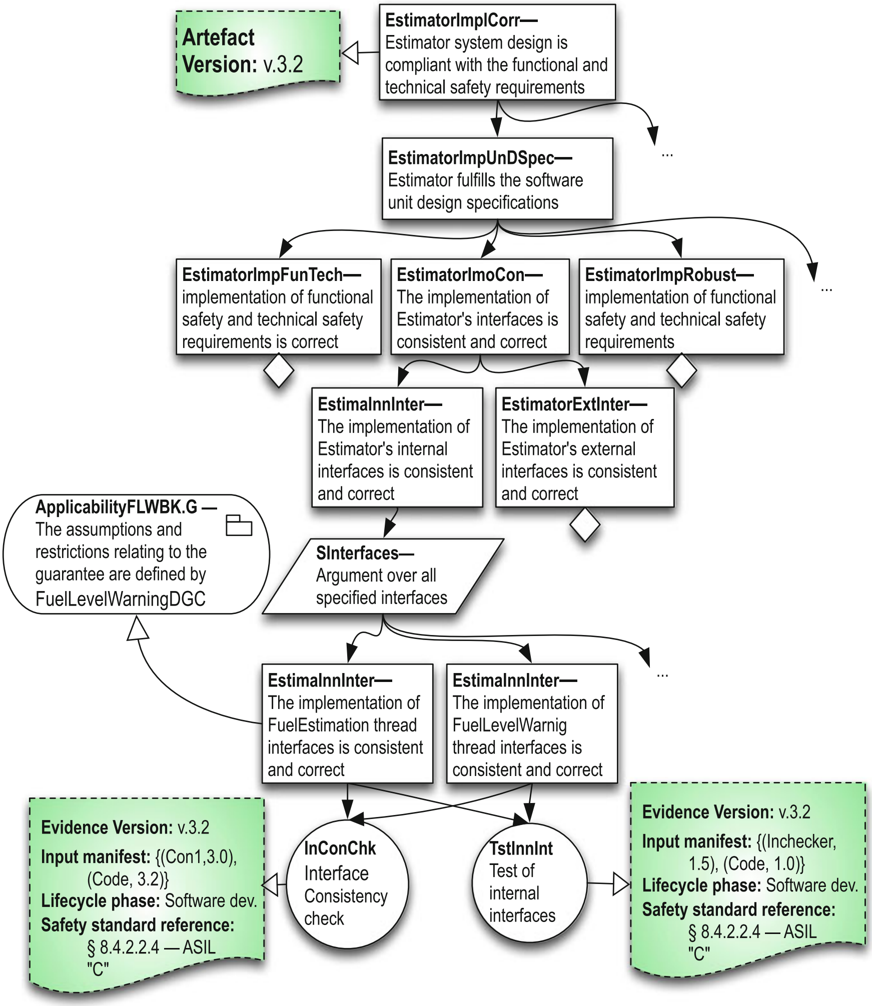

Artefact Version: v.3.2
EstimatorImplCorr— Estimator system design is compliant with the functional and technical safety requirements
...
EstimatorImpUnDSpec— Estimator fulfills the software unit design specifications
EstimatorImpFunTech— implementation of functional safety and technical safety requirements is correct
EstimatorImoCon— The implementation of Estimator's interfaces is consistent and correct
EstimatorImpRobust— implementation of functional safety and technical safety requirements
...
EstimaInnInter— The implementation of Estimator's internal interfaces is consistent and correct
EstimatorExtInter— The implementation of Estimator's external interfaces is consistent and correct
ApplicabilityFLWBK.G — The assumptions and restrictions relating to the guarantee are defined by FuelLevelWarningDGC
SInterfaces— Argument over all specified interfaces
EstimaInnInter— The implementation of FuelEstimation thread interfaces is consistent and correct
EstimaInnInter— The implementation of FuelLevelWarnig thread interfaces is consistent and correct
...
Evidence Version: v.3.2
Input manifest: {(Con1,3.0), (Code, 3.2)}
Lifecycle phase: Software dev.
Safety standard reference: § 8.4.2.2.4 — ASIL "C"
InConChk Interface Consistency check
TstInnInt Test of internal interfaces
Evidence Version: v.3.2
Input manifest: {(Inchecker, 1.5), (Code, 1.0)}
Lifecycle phase: Software dev.
Safety standard reference: § 8.4.2.2.4 — ASIL "C"

Fig. 11 An argument fragment of SW Integration test safety case module

no goals are supported by contracts. The integration, in our example, is done through public and away goals.

4.2.4 Assess/Improve Change Impact

In this step, we use our approach for maintaining safety cases (in Sect. 2.4) to extend IAWG's DGC. We use the extended DGC in the FLES example to show how the extension can help: (1) highlighting the affected argument elements, and

Table 4 FuelLevelWarningDGC

<table>
<tr><td colspan="6">Dependency—guarantee contract | FuelLevelWarningDGC</td></tr>
<tr><td>Consumer Dependency</td><td>Integrator</td><td>Provider Guarantee</td><td colspan="3">Artefact Version</td></tr>
<tr><td><b><i>FuelLevelWarningBK.G1</i></b></td><td>Supported by away goal <i>FuelEstimationBK.G5</i></td><td><b><i>FuelEstimationBK.G5</i></b></td><td colspan="3">V.3.2</td></tr>
<tr><td><i>totalFuelLevel</i> value is received</td><td>Is Supported By</td><td>Provides the <i>totalFuelLevel</i> value</td><td colspan="3"></td></tr>
<tr><td><i>totalFuelLevel</i> value is received via <i>GetEstimatedFuelLevelValue_2 port</i></td><td>Is Consistent with</td><td>The <i>totalFuelLevel</i> value is sent on port <i>SetSensorValue</i>.</td><td colspan="3">InConChk TstInnInt</td></tr>
<tr><td><i>totalFuelLevel</i> data format is defined by FLES {Interface Specification Ref.20}</td><td>Is Consistent with</td><td><i>totalFuelLevel</i> data format is defined by FLES {Interface Specification Ref.20}</td><td colspan="3"></td></tr>
<tr><td colspan="6">Supporting evidence</td></tr>
<tr><td>No</td><td>GSN element</td><td>Evidence version</td><td>Input manifest</td><td>Lifecycle phase</td><td>Safety standard reference</td></tr>
<tr><td><b>1</b></td><td>InConChk</td><td>V.3.2</td><td>(Inchecker, 1.5), (Code, 1.0)</td><td>SW Dev.</td><td>§ 8.4.2.2.4 ASIL “C”</td></tr>
<tr><td><b>2</b></td><td>TstInnInt</td><td>V.3.2</td><td>(Con1, 3.0), (Code, 3.2)</td><td>SW Dev.</td><td>§ 8.4.2.2.4 ASIL “C”</td></tr>
</table>

(2) identifying inadequacies in the generated artefacts from the development lifecycle of FLES.

Table 4 shows an extended DGC of *FuelLevelWarning.BK* The extension is represented by the cells in grey. Moreover, Fig. 11 shows items of evidence (i.e., GSN solution) that support claims about the consistency among the ports of FLES blocks. The green elements in the figure represent the annotations described in Sect. 2.4.

Now, let us consider the potential change scenario in Sect. 4.2.1 to illustrate how the information contained within the annotations aids the change impact analysis in safety arguments. Merging *FuelEstimation* and *FuelLevelWarning* into one thread will impact the consistency of the interfaces and connections of FLES. Suppose that an engineer making this change had updated the artefact version annotation(s) in part of the argument that refers to the interfaces of those threads. An automated implementation of the described checks in Sect. 2.4 could highlight the need to re-run the interface consistency check, as well as, the *Estimator* internal interfaces testing. If the new version of the implementation is version 3.3, analysis of the manifest associated with *InConChk* and *TstInnInt* would reveal evidence based on an older version of the implementation and tools could flag *InConChk* and *TstInnInt* as out-of-date and suspect. Automated analysis might also highlight goal *EstimatorImpCorr* because its artefact version annotation refers to an out-of-date version of the Estimator implementation. The goal and its supporting argument are suspect because they might refer to parts of the implementation that no longer exist or make claims about the implementation that are no longer true.

Table 5 shows the impacted elements of the safety case with a brief explanation for each element.

The principal difference between our work and the existing approach proposed by the IAWG MSSC is that the MSSC approach contains changes at the level of a safety argument module and the corresponding system blocks. In contrast, our approach provides the engineer to contain the changes at a lower-level where they feel that a tighter control over change is needed. More specifically, our approach means that changes can be contained within a safety argument module and within

Table 5 Results of change impact analysis

No.	Module name	Element affected	Explanation
1	*SW Safety Requirements*	DecF1010TSR1	The decomposition of this requirement has been changed
2	*SW Safety Requirements*	DecF2010TSR1	The decomposition of this requirement has been changed
3	*FuelLevelWarning. BK*	The entire module	Merged with another Module
4	*SW Integration test*	EstimaInnInter and all claims below	Argument about the estimation internal interfaces is suspect
5	*SW Integration test*	*InConChk*	Out of date implementation
6	*SW Integration test*	*TstInnInt*	Out of date implementation

specific system blocks. It could be argued that this could have been handled in the existing approach by decomposing the system and its safety argument differently, however in practice it is better not to constrain system architects unnecessarily.

5 Conclusion and Future Work

Applying changes to systems during their lifetime is inevitable task. In safety critical systems, system changes can be accompanied with changes to safety arguments. Maintaining those arguments is painstaking process because of the dependencies between their elements. The IAWG MSSC process was introduced as a response to safety cases maintenance difficulties. The process recommends applying changes as a series of relatively small increments rather than occasional major ones. However, The guidance of MSSC process does not show detailed information about how to follow some steps including the impact analysis part. In this paper, we applied the process to a real safety critical system to show how system engineers can identify the elements in a safety argument that might be impacted by a change. We showed that by extending the proposed DGC by IAWG to include additional information as annotations that is useful to highlight the impacted argument elements. Moreover, we provided starting points to maintain the affected parts of the argument as we described the reasons why they have become inadequate due to the change. The impact check based on the additional information is still manual as we have not yet studied the feasibility or value of developing a tool to automate the checks but we leave this effort to future work.

Acknowledgments We acknowledge the Swedish Foundation for Strategic Research (SSF) SYNOPSIS Project for supporting this work.

References

1. Jaradat O, Graydon PJ, Bate I (2014) An approach to maintaining safety case evidence after a system change. In: Proceedings of the 10th european dependable computing conference
2. EUROCONTROL European organisation for the safety of air navigation, preliminary safety case for enhanced traffic situational awareness during flight operations, PSC ATSA-AIRB. www.eurocontrol.int/articles/cascade-documents. Accessed 20 Feb 2015
3. Ewan D, Whiteside I (2012) Hierarchical safety cases. Technical report NASA/TM-2012-216481, NASA Ames Research Center
4. Kelly T, McDermid J (1999) A systematic approach to safety case maintenance. In: Felici M, Kanoun K (eds) Computer safety, reliability and security, vol 1698., Lecture Notes in Computer ScienceBerlin, Springer, pp 13–26
5. Conmy P (2005) Safety analysis of computer resource management software. Ph.D. thesis, University of York. https://www.cs.york.ac.uk/ftpdir/reports/2006/YCST/07/YCST-2006-07.pdf. Accessed 5 Mar 2015

6. Kelly T (2007) Modular certification. Lecture Note. http://webhost.laas.fr/TSF/IFIPWG/Workshops&Meetings/52/workshop/10%20Kelly.pdf. Accessed 20 Feb 2015
7. IAWG MSSC Process (2012) Modular Software Safety Case Process Description. https://www.amsderisc.com/wp-content/uploads/2013/01/MSSC_201_Issue_01_PD_2012_11_17.pdf. Accessed 20 Feb 2015
8. Kelly T (2006) Using software architecture techniques to support the modular certification of safety-critical systems. In: Eleventh Australian workshop on safety critical systems and software, Australia
9. ISO 26262 (2011) Road vehicles—functional safety. International organization for standardization
10. Origin Consulting (2011) GSN Community Standard. http://www.goalstructuringnotation.info/. Accessed 20 Feb 2015
11. Kelly T (1995) Literature survey for work on evolvable safety cases. Department of Computer Science, University of York
12. Wilson SP, Kelly TP, McDermid JA (1997) Safety case development: current practice, future prospects. In: Proceedings of software bases systems—12th annual CSR workshop
13. Fenn JL, Hawkins RD, Williams P, Kelly TP, Banner MG, Oakshott Y (2007) The who, where, how, why and when of modular and incremental certification. In: Proceedings of the 2nd IET international conference on system safety, pp 135–140
14. Jaradat O, Graydon P, Bate I (2013) The role of architectural model checking in conducting preliminary safety assessment. In: Proceedings of the 31st international system safety conference
15. Jaradat O (2012) automated architecture-based verification of safety-critical systems. Master thesis. Mälardalen University, Sweden. www.diva-portal.org/smash/record.jsf?pid=diva2%3A723310&dswid=5193, Accessed: 20 Feb 2015

Track Maintenance Between Trains Simulation

Christer Stenström, Ulla Juntti, Aditya Parida and Uday Kumar

Abstract Infrastructure managers (IMs) need to plan their maintenance work about 1½–2 years before a new train time table (TTT) comes into force to minimise the effect on traffic. Maintenance work that is planned in less than 1 year ahead of the TTT has to compete with, or need to be fitted into, operators' applications for capacity. However, maintenance work is at times planned only a few weeks before execution, and depending on the railway line in question, a few hours during night can be available for maintenance. In addition, sudden failures in track normally require repair immediately or within a day. If rail transportation increases, it also becomes harder to find time in track for maintenance. Therefore, it is of interest to simulate maintenance tasks between trains to minimise track maintenance possession time. Such simulation can be used to: study maintenance work in TTTs with random and regular train departures; study the effect of exceeding allocated maintenance windows; and to study the effect of increase in train frequency. In this paper, Monte Carlo method is applied to simulate track maintenance between trains as a function of train frequency.

Keywords Maintenance · Operation · Rail infrastructure · Maintainability · Maintenance supportability · Planning · Scheduling · Maintenance window · Service window · Monte Carlo method

C. Stenström (✉) · A. Parida · U. Kumar
Division of Operation and Maintenance Engineering, Luleå University of Technology,
971 87 Luleå, Sweden
e-mail: christer.stenstrom@ltu.se

A. Parida
e-mail: aditya.parida@ltu.se

U. Kumar
e-mail: uday.kumar@ltu.se

U. Juntti
Performance in Cold AB, Storgatan 11, 972 38 Luleå, Sweden
e-mail: ulla.juntti@minus8.nu

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_26

1 Introduction

The aim of rail infrastructure maintenance planning is to minimise the effect on traffic. Therefore, infrastructure managers (IMs) start planning their maintenance work far ahead of time to fit with the planning of train time tables. In the European Union (EU), each IM is obligated to publish a Network Statement (NS) with all the information needed for train operators interested in capacity [1]. The NS must be published about a year before a new train time table comes into force. Consequently, the planning of major maintenance work starts about 1½–2 years before a new train time table. Maintenance work that is planned after submission of the NS has to compete with, or be fitted into, operators' applications for capacity. However, some work is planned only a few weeks before execution, and depending on the railway line in question, a few hours during night can be available for maintenance. Besides late planned maintenance work, failures in track normally require repair immediately or within 1 day. In addition, with increasing rail transportation it becomes harder to find time in track for maintenance [2]. Therefore, it is of interest to simulate maintenance tasks between trains. Such simulation can be used to:

- Study maintenance work in train time tables with random train passages to find suitable time windows, e.g. freight trains.
- Study maintenance work in train time tables with regular train passages/departures, i.e. urban areas. As an example, a certain maintenance work can take a very long time to complete at a train frequency of 10 min, while cancelling a group of trains or running one train out of three may be impacting the train service unnecessary negatively; i.e. cancelling each second train may then be the most optimal choice.
- Study the effect of exceeding allocated maintenance windows. As an example, the available maintenance window may be 4 h, but it is known from experience and historical data that the work in question takes 5–6 h. Possible solutions are: carry out the work in one shift, exceeding the 4 h, and do the last work between the trains; do the work in two shifts, i.e. 8 h over 2 days; or increase the number of personnel in the maintenance team if possible.
- Study the effect on maintenance of future increases in the frequency of trains. This is especially important in maintenance contracting, as contracts are often performance based and stretch over several years [3]. If a possible increase in the numbers of trains running is not properly taken care of within the maintenance contracts, the infrastructure manager and contractor may end up in a disagreement.

The study of maintenance between trains was initiated in the AUTOMAIN project [4, 5]. In AUTOMAIN, the effect of exceeding an allocated maintenance window was studied by comparing an alternative maintenance approach.

Specifically, the simulation concerned the decision to use one or two welding teams for frog (common crossing) replacement of switches and crossings (S&Cs). Maintenance work between regular train departures, i.e. urban areas, was also studied in an attempt to balance maintenance cost and train services. As a continuation of the study in AUTOMAIN, Monte Carlo method is applied to simulate track maintenance between trains as a function of train frequency.

2 Methodology

Maintenance time in terms of travel, preparation, repair and clearance can be estimated through experience and use of historical work order data. However, the actual time to complete a particular maintenance task depends on the train time table and safety regulations governing entrance to and closure of the railway section in question. With these inputs, the actual time to maintenance can be estimated. Matlab software is used for model construction and Monte Carlo simulation. Monte Carlo method is used to predict increases in maintenance time by sampling random train time tables (TTTs). The model has a number of input and output parameters. The input parameters are as follows:

- Train time table (TTT)
- Non-value adding (NVA) time (t_{NVA}): Consist of preparation, confirmation, communication, waiting and lost time, for entrance to and closure of track
- Active repair time (t_{Active})
- Minimum time for maintenance (t_{Min})
- Arrival point in the time table

Minimum time for maintenance (t_{Min}) can be set as a fixed value or based on t_{Active}. As an example, if the required time for a maintenance activity is 150 min and t_{Min} is set to 10 % of that time, i.e. 15 min, then no maintenance will be carried out if the time left for maintenance between two trains is less than 15 min.

The output parameter is the (actual) maintenance time: the time from the arrival of the maintenance team until the work is finished and the track is cleared.

Figure 1 demonstrates the model. Given that t_{NVA} equals 10 min, t_{Active} equals 50 min and $t_{Min} = 0.1t_{Active} = 5$ min, the maintenance time becomes 109 min, i.e. 118 % more than the 50 active minutes required.

Random TTTs are generated using a uniform distribution, with the exception that adjacent trains must have a minimum distance in time of x minutes. An example of a 120 min maintenance task in randomly drawn TTTs is shown in Fig. 2.

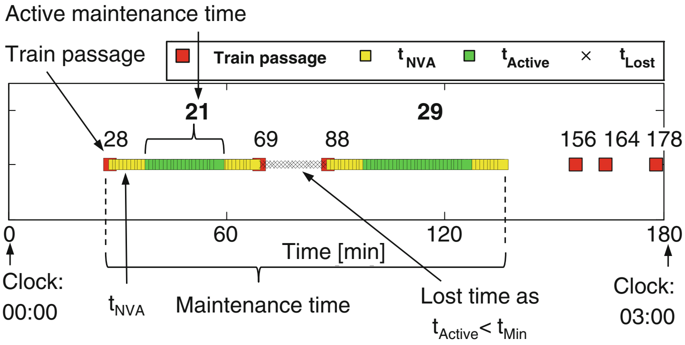


Fig. 1 Demonstration of the model

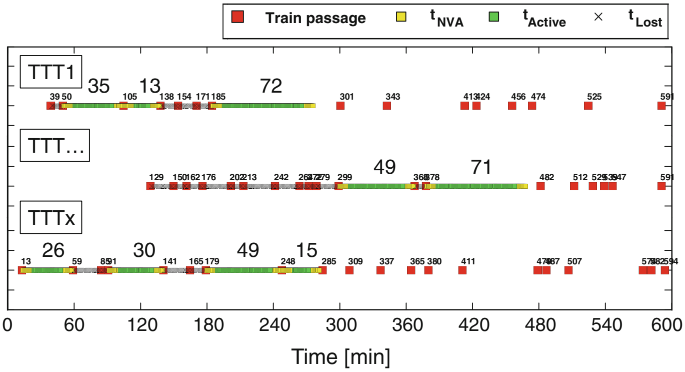


Fig. 2 Example of random drawn TTTs with a 120 min maintenance work

3 Results

The number of trains occupying the Iron Ore Line between Kiruna, Sweden, and Narvik, Norway, is predicted to increase from 42 trains per day in 2014 to 50 trains per day in 2020 [6]; see Fig. 3.

For the simulation: operating time is set to 18 h per day; the number of trains per day is set to range from 10 to 60; a random uniform distribution is used to set out trains passages; an exception is added whereby adjacent trains must have a minimum distance in time of 5 min; t_{NVA} is set to 5 and 10 min, giving two different cases; t_{Active} equals 120 min; t_{min} equals 10 min; and the number of random TTTs is set to 1000 for each train frequency and t_{NVA}. The result is shown in Fig. 4: the data points are the mean values; their whiskers are the standard deviations; and the density function and box plot, with whiskers of 1.5 IQR (interquartile range), are for the highest data point. As indicated in the figure, the maintenance time has an exponential or polynomial growth as the train frequency increases.

Fig. 3 Predicted increase in train frequency of the iron ore line [6]

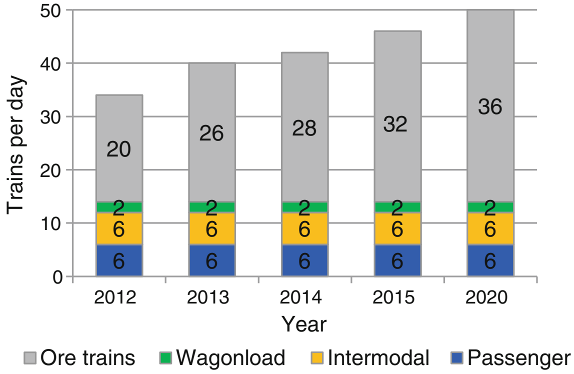
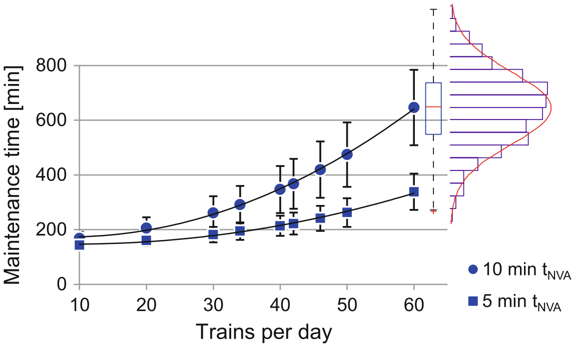


Fig. 4 Predicted increase in maintenance time with increase in train frequency



By comparing the results with the predicted increase in trains, it is found that the average maintenance time increases with 29 % (t_{NVA} = 10 min) when the number of trains per day increases from 42 to 50; see Fig. 5. With t_{NVA} of 5 min, the maintenance time increases with 18 %.

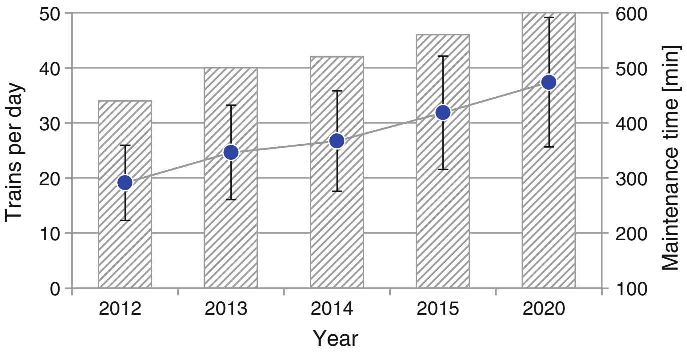


Fig. 5 Predicted train frequency and maintenance time

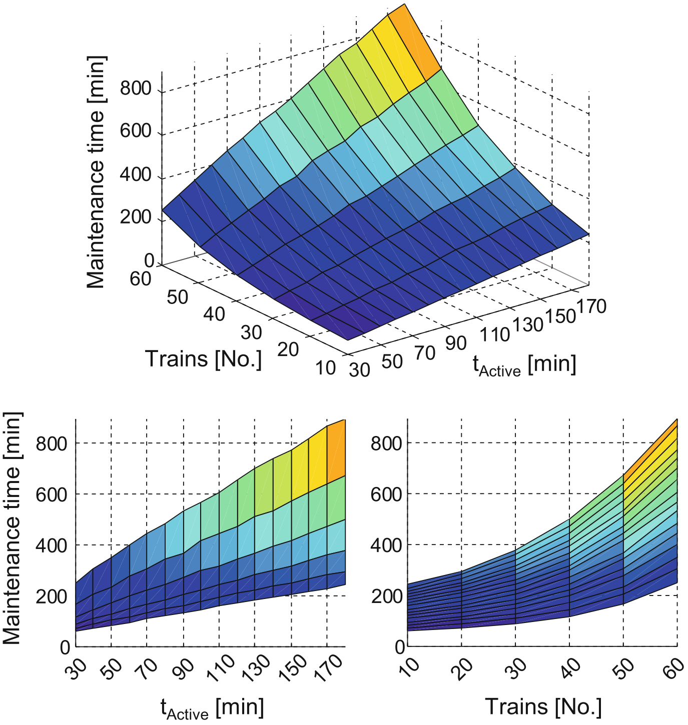


Fig. 6 Maintenance time as a function of trains and active repair time (t_{Active})

In the results above, t_{Active} equals 120 min. If t_{Active} is set to vary, while keeping the other constants as given above, with t_{NVA} = 10 min, it is seen in Fig. 6 that the maintenance time increases linearly with t_{Active}. Each data point in Fig. 6 is the mean value of 1000 simulations.

4 Discussion

The method for predicting increase in maintenance time, as a function of increase in train frequency, is based on random drawn TTTs with uniform distributions. However, depending on the railway line in question, the configuration of a future TTT, with more trains, can be known. It is still not known, however, when a failure will take place, i.e. randomness. Thus, the presented method can alternatively be used with a fixed TTT and some distribution for failures. Sampling TTTs or sampling failures will, to a certain extent, yield similar results. As an example, randomly setting out a failure in a TTT is analogous to randomly setting a TTT to a failure.

The simulation considered a maintenance work that can be temporarily stopped to let trains pass. Other maintenance work can include work steps that have to be finished before trains can pass. As an example, frog replacement requires two out of four welds to be completed before trains can pass. This type of maintenance must be put into allocated maintenance windows, e.g. at night, or in sufficiently large windows within the TTT. This kind of works can be simulated by sampling random TTTs with a fixed maintenance window. Thus, by including the various types of maintenance work, a final predicted increase in maintenance time can be achieved. Historical maintenance data can be used for this purpose, together with expert judgement. It should also be noted that adding the logistic time (travel time) will reduce the increase in maintenance time as the logistic time is unaffected by the train frequency.

5 Conclusions

Monte Carlo method can be used to predict maintenance time as a function of train frequency. It has been shown that the maintenance time increases exponentially with train frequency (Fig. 4). This effect is due to safety regulations governing entrance to and closure of railways.

Specific to the case study, with the model input values used, the average maintenance time on the Iron Ore Line in Sweden, will increase by $\sim 30\ \%$ ($t_{NVA} = 10$ min) from year 2015 to 2020. Nevertheless, expert opinion and parameter study can improve the prediction, and is required if the method is applied in practice.

Acknowledgments This work was supported by Luleå Railway Research Centre (JVTC) and Trafikverket (Swedish Transport Administration). The authors would also like to thank Prof. Per-Olof Larsson-Kråik for his intellectual support in the project.

References

1. EC (2001) Council directive 2001/14/EC of the European Parliament and of the council of 26 February 2001 on the allocation of railway infrastructure capacity and the levying of charges for the use of railway infrastructure and safety certification. Official J Eur Commun (L75), pp 29–46
2. EC (2011) White paper: roadmap to a single European transport area—towards a competitive and resource efficient transport system. In: COM (2011) 144. European Commission (EC), Brussels
3. Famurewa SM, Asplund M, Galar D, Kumar U (2013) Implementation of performance based maintenance contracting in railway industries. Int J Syst Assur Eng Manag 4(3):231–240
4. Juntti U, Parida A, Stenström C, Famurewa SM, Asplund M, Nissen A, Ripke B, Lundwall B (2013) Optimised maintenance activities like, grinding, tamping and other maintenance processes. D4.2. AUTOMAIN project

5. Parida A, Juntti U, Stenström C, Famurewa SM (2014) Capacity enhancement through optimized maintenance of railway networks. In: EuroMaintenance 2014 congress proceedings, 5–7 May 2014, pp 409–416, EFNMS (European Federation of National Maintenance Societies)
6. Boysen H (2013) Quicker meets, heavier loads and faster empties—effects on transportation capacity and cycle time. In: 10th international heavy haul association (IHHA) conference 2013, IHHA

Process Analysis of Human Failures in Railway Maintenance

Mattias Holmgren and Peter Söderholm

Abstract The aim of maintenance is to retain and restore the required functions of technical systems to ensure the productivity and efficiency of technical systems. However, improper maintenance, in the sense of incorrectly performed maintenance or lack of suitable maintenance activities, contributes to deterioration and, even more seriously, cause incidents and accidents. Therefore, it is important to identify hazards from occurred incidents and accidents to learn and avoid these hazards in the future. For that matter, in turn, documentation from occurred events should be done in a systematic way and then be an important part of the management system—and last, but not least, used for continuous improvement and risk reduction. This study aims to describe a process-oriented approach to analyse causes of human failures contributory to maintenance-related incidents and accidents, in order to support continuous risk reduction. The proposed methodology with supporting tools can be used for analysis purposes and guide decision making. The approach is illustrated by a case study at the Swedish railways.

Keywords Maintenance · Risk · Process · Railway · Human failure · Sweden

1 Introduction

In order to achieve continuous improvement and risk reduction, a process view is central. When considering the maintenance process, it should be vertically aligned with the organisation's overarching goals in order to be effective, i.e. to do the right things. In addition, the maintenance process should also be horizontally aligned

M. Holmgren (✉)
Luleå University of Technology (LTU), Universitetsvägen 1, 971 87 Luleå, Sweden
e-mail: mattias.holmgren@ltu.se

P. Söderholm
Trafikverket (Swedish Transport Administration), LTU, Box 807, 971 25 Luleå, Sweden
e-mail: peter.soderholm@trafikverket.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_27

with the operational and modification processes to be efficient, i.e. to do things the right way [1–4].

Furthermore, the maintenance process and its included phases should be related to appropriate methodologies and tools in order to support core values of the organisation. Examples of core values that the maintenance process should support are base decisions on fact, improve continuously, and focus on processes [2, 4–6].

Maintenance is intended to retain and restore the required function of technical systems to deliver desired services. However, there are multiple examples when maintenance has failed to achieve this and also contributed to extensive losses. One safety-critical application area that experience the unwanted contribution of maintenance is railway. Two examples of this are the derailments at Ladbroke Grove and Hatfield in the United Kingdom [7].

In October 1999, a major derailment and collision occurred at Ladbroke Grove. As a result of the collision 31 people died and 227 were taken to hospital. Two major conclusions were drawn. Firstly, the process for the judgement of contracts was not operated with due regard for training and preparation of the contract workforce. Secondly, the managerial control of the work performed by maintenance contractors and sub-contractors was inadequate. In October 2000, four people were killed in a derailment near Hatfield. The accident investigation showed that the immediate cause of the derailment was a fragmentation of the rail caused by neglected maintenance actions [7].

Two later examples from the United Kingdom are from 2007 and 2008. In April 2007, a track welder was struck by a train and fatally injured at Ruscombe Junction. The local practice was that the track work was carried out within the safety zone of the track, called “The Red Zone”. The track welder continued to work on the site although he had been warned for the approaching train [8].

In December 2008, a track worker was struck by a train and injured at Stevenage, Hertfordshire, in the United Kingdom. The track worker moved out of the position of safety to a point where he could come into contact with the train. The planning process was insufficiently detailed to identify all the hazards and adequate safe systems of work [9].

In Sweden there are also multiple examples of maintenance related accidents, e.g. Kimstad in 2010 where a maintenance vehicle collided with a train, which resulted in one killed and 20 injured persons [10, 11].

One Swedish study showed that maintenance-related causes represent 30 % of all rail and track related incidents and accidents during the time period of 1988–2000. At the same time, maintenance-related risks are often manifested in maintenance execution (i.e. the sharp end), while its causes may be located somewhere else (i.e. the blunt end). The case study showed that about 80 % of the maintenance-related accidents in Swedish railway happen during the execution phase. In addition, maintenance-related activities are often scattered and can be found within other processes and areas of the organisation [12].

Holmgren also discovered that the most common cause of maintenance-related accidents is imperfect communication and information between the maintenance personnel and the operators. This cause was followed by rule violations, especially lack of permission to perform maintenance work on the track, as the second most frequent cause [12].

In another Swedish study of unwanted events (accidents, incidents, or deviations contributing to risk) related to work in track during the time period of 2007–2012, 72 % were classified as caused by human failure [13].

In the investigations related to the maintenance-related accidents cited above, one common cause is human failure. However, human error as an explanation for accidents is unsatisfactory, since there are always organisational and operational aspects that lay the foundation for these errors [14]. Hence, errors are consequences, rather than causes [15]. In other words, human errors are the result of a network of actions and conditions which involve people, teams, tasks, workplace and organisational factors [16]. Hence, discovering a human error is the beginning of the search for causes, not the end [15, 17]. The intention should be to identify and control hazardous conditions, instead of focusing on single causes of accidents and trying to eliminate them [18].

Based on the problem description above, the purpose of this study is to describe a process-oriented approach for identification of causes of human failures related to maintenance, in order to support continuous improvement and risk reduction.

2 Method and Material

The overall research strategy applied in this study is a single case study at Trafikverket (Swedish Transport Administration), with focus on railway maintenance and related risks.

Two complementary sources of secondary data have been used to illustrate the proposed process approach. The first source of data, is based on Trafikverket's data base with accidents, incidents and deviations contributing to risk within work environment, traffic safety, power safety, protection against crime and fire (and to some extent also quality, environment, and audits). The selection criteria was unwanted events related to engineering works during 2007–2012, which resulted in a total of 703 unwanted events (262 accidents, 358 incidents, 83 risks). This data is originally structured according to the logic of MTOY-analysis (man, technology, organisation, external events). Some of this material is reported by [13].

The second source of data is 26 maintenance-related accident investigations (for events occurred between the years 1988–2000), where human failure has been identified as a contributory cause. The average number of pages for these analysed investigations is 30, varying between 11 and 154 pages. Over the years, these investigations tend to follow the MTOY-approach to an increasing degree. The background and rationale for this data is described in more detail by [19].

The two sources of secondary data have then been used in a qualitative analysis based on a generic maintenance process. Hence, especially in the analysis of the first source of data, the process-approach has been used in combination with the MTOY-methodology. For the second source of data, the process-analysis approach has been performed in combination with HAZOP (Hazard and operability studies) guidewords as a supporting tool.

2.1 Case Study Description from a Risk Management Perspective

On an aggregated level, Trafikverket applies an integrated enterprise risk management approach that combines the areas of risk management (ISO 31000) [20], incident, continuity, and crisis management (e.g. according to the series of ISO 22300 [21], societal security, and 27000, information security [22]), where the organisation's capability within these areas is judged systematically through assessment activities influenced by quality management (ISO 9000) [23].

In addition, there are specific areas related to risk management that are integrated through the overall framework, e.g. information security (ISO 27000) [22], quality management (ISO 9000) [23], and asset management (ISO 55000) [24]. When the risks are related to traffic safety within railway, the approach described in the EU-directive Common safety methodology (SCM)—Risk assessment and analysis (RA) should be applied.

One cornerstone of the integrated enterprise risk management approach applied at Trafikverket is to identify risks in the processes that can affect the capability to deliver the desired level of quality. It is also desired that both the risks and related control activities to manage the risks should be highlighted within the process descriptions. For this purpose, the proposed process approach is believed to be a useful methodology.

2.2 Process and Risk Approach to Maintenance

Maintenance activities can be seen as following the IEC suite of dependability management standards [25, 27], which means that the maintenance process can be described as in Fig. 1.

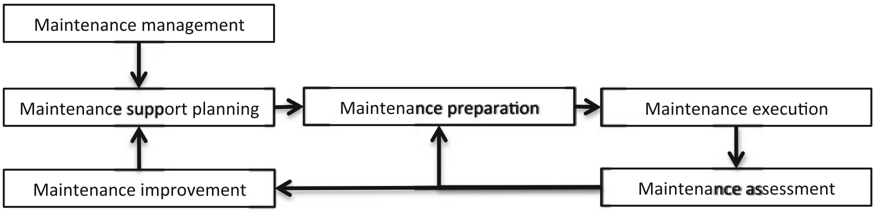


Fig. 1 Generic process model of maintenance. In accordance with IEC 60300-3-14

The activities that are related to maintenance depends on their purpose. Hence, if the purpose is to maintain or retain the required function in order to deliver a desired service, it is classified as a maintenance-related activity. The activities can in turn be related to any of the phases of the maintenance process, i.e. maintenance management, maintenance support planning, maintenance preparation, maintenance execution, maintenance assessment, and maintenance improvement.

These phases can also be related to specific methodologies and tools. Hence, the phase of maintenance management is closely related to the ISO 55000 standard [24], and the risk-related part of this phase is mainly related to ISO 31000 [20] and COSO ERM (Committee of sponsoring organizations of the Treadway commission —Enterprise risk management) [26]. The phase of maintenance support planning is closely related to different technical systems within the railway (e.g. signalling, catenary, telecommunication, and track systems), where risk-related methodologies such as Reliability-centred maintenance (RCM) [27] and Failure, mode, effects, and criticality analysis (FMECA) [28] are applied. The two phases of maintenance planning and execution are on the other hand closely related to the technical systems installation on specific lines, e.g. through considering continuity management (ISO 22300) [21] to achieve the desired quality of service level (where, besides safety, punctuality and regularity normally are the most important measures). The assessment and improvement phases of the maintenance process are in turn closely related to methodologies within the quality management area, e.g. as described in the ISO 9000-series [23]. In fact, one important part of viewing maintenance as a process with the included phases is to support the work with continuous improvement and risk reduction through similarities with the improvement cycle as described by [29, 30].

2.3 Analysis of Human Failures in Maintenance

Human failures are seen as consisting of both human errors and rule violations. Human error is in turn occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome, and when these failures cannot be attributed to the intervention of some chance agency [31].

Rule violation is seen as deviations from safe and established procedures, standards or rules to control a system. Hence, rule violation may be either deliberate or erroneous [32].

In this paper, rule violations are seen as deliberate actions, even though the outcome is unintended. If the outcome is intended, the human action is classified as sabotage.

A system approach to analyse human failure in combination with other causes is the MTO-analysis, where human, organizational, and technical factors should be focused equally in an accident investigation. The methodology is based on HPSES (Human performance enhancement system) from the nuclear industry. The version

Table 1 A classification of human failures akin to hazard and operability studies (HAZOP) guidewords

Action errors
• A1 Operation too long/short • A2 Operation mistimed • A3 Operation in wrong direction • A4 Operation too little/much • A Operation too fast/slow • A6 Misalign • A7 Right operation on wrong object • A8 Wrong operation on right object • A9 Operation omitted • A10 Operation incomplete • A11 Operation too early/late
Checking errors
• C1 Check omitted • C2 Check incomplete • C3 Right check on wrong object • C4 Wrong check on right object • C5 Check too early/late
Information retrieval errors
• R1 Information not obtained • R2 Wrong information obtained • R3 Information retrieval incomplete • R4 Information incorrectly interpreted
Information communication errors
• I1 Information not communicated • I2 Wrong information communicated • I3 Information communication incomplete • I4 Information communication unclear
Selection errors
• S1 Selection omitted • S2 Wrong selection made
Planning errors
• P1 Plan omitted • P2 Plan incorrect
Violations
• V1 Deliberate actions

applied at Trafikverket also considers external factors, e.g. climate, weather, and third party interventions. For further description of MTO-analysis, see [33, 34]. One way to classify human failures is listed in Table 1 [35].

3 Results

In this section the result of the two process-analyses are presented.

3.1 Process-Analysis Combined with MTO-Analysis

The MTO-analysis is based on a system perspective. In Fig. 2, it is seen that the majority of causes are classified as human failure according to the MTO-analysis. However, when considering human failures from a process perspective, all causes of the MTO-analysis can be structured according to the process approach and also be viewed from an actor perspective within the different phases of the process.

Since the unwanted events that have been studied focuses on engineering works, i.e. work in track, the human failure classification (Fig. 3) can mainly be related to the maintenance execution phase. However, since railway is highly regulated the part of human failure classified as violation can also indicate possible improvements

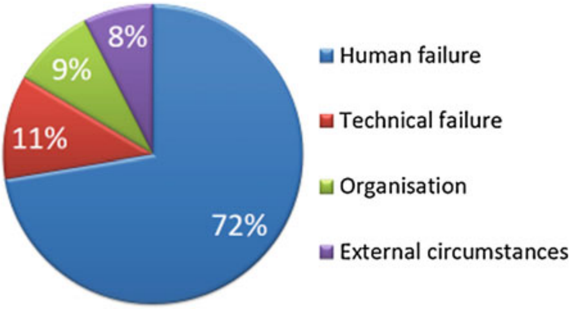


Fig. 2 Causes of unwanted events related to track work 2007–2012 in Sweden

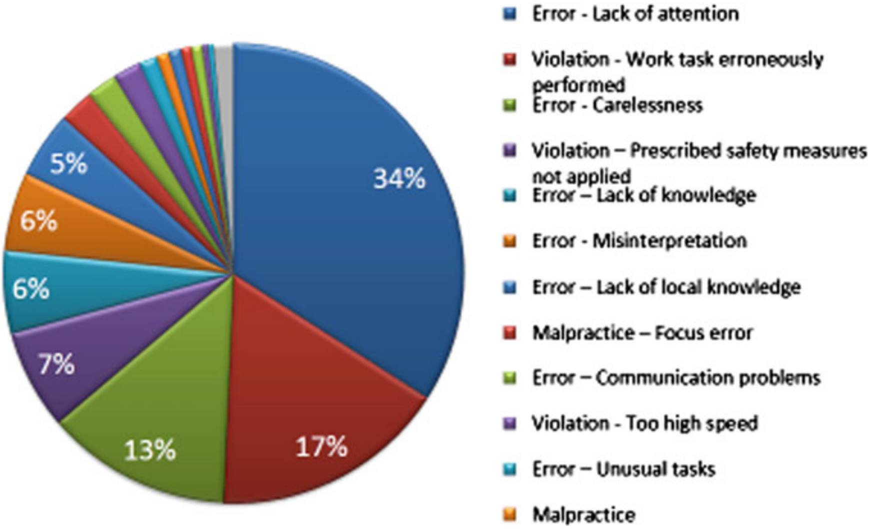


Fig. 3 Causes of human failure—unwanted events related to track work 2007–2012 in Sweden

in the regulations, e.g. at the maintenance support planning phase. The violations can also indicate improvement possibilities at the maintenance preparation phase, e.g. an improved planning of maintenance execution that actually allows the personnel to follow the regulations. A complementary approach to classify these human failures would be to apply the HAZOP guidewords.

The organizational causes (Fig. 4) can mainly be more directly related to other phases than execution of the maintenance process, e.g. maintenance support planning (e.g. regulation) and maintenance preparation (e.g. planning activities).

Technical causes of unwanted events that are related to working vehicles, equipment and tools indicate possibilities to improve the capability of the maintenance process, e.g. mainly during the preparation and execution phases. In addition, the technical causes related to the infrastructure may indicate possibilities to improve maintenance practice, e.g. through changes of type and frequency of inspection in the phase of maintenance support planning. The technical causes related to infrastructure may also be addressed by modifications of the technical system, however, this is not (by definition) any maintenance activity. See Fig. 5.

The external causes of unwanted events related to track work mainly seem to address changes of the technical system to improve its robustness, which are not related to maintenance but rather to modification (see Fig. 6). However, the external causes can also indicate capabilities within the process phases that should be strengthened in order to achieve an increased robustness, e.g. proactivity through risk and security measures, but also improved response times through incident, continuity and crisis management. These activities related to different aspects of integrated risk management can mainly be addressed within the phases of maintenance support planning, maintenance preparation, and maintenance execution (cf. Sect. 2.2—Process and risk approach to maintenance).

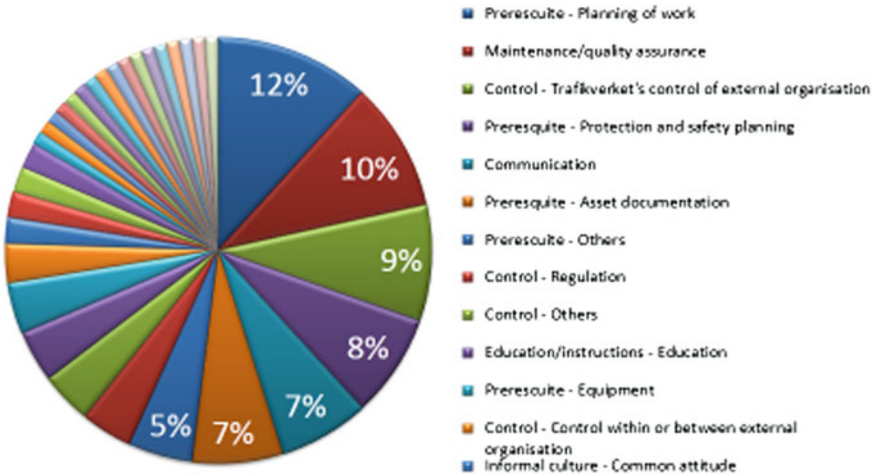


Fig. 4 Causes of organisational failure—unwanted events related to track work 2007–2012 in Sweden

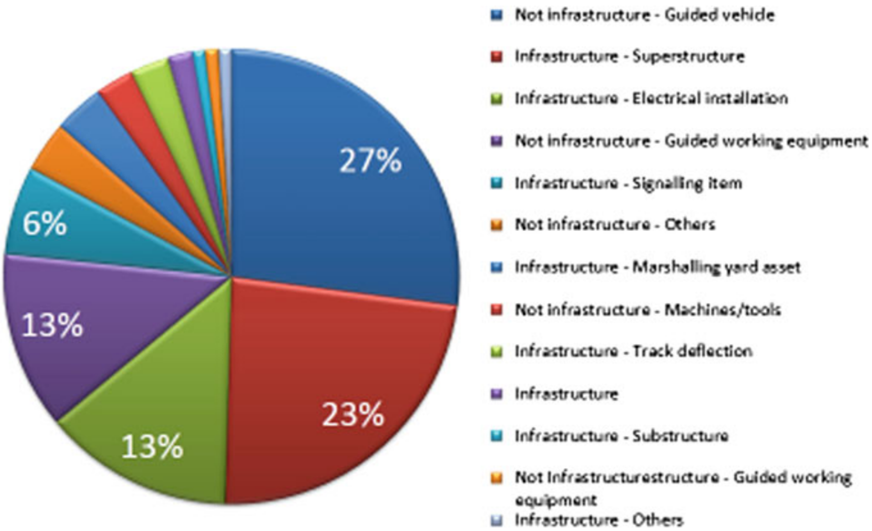


Fig. 5 Causes of technical failure—unwanted events related to track work 2007–2012 in Sweden

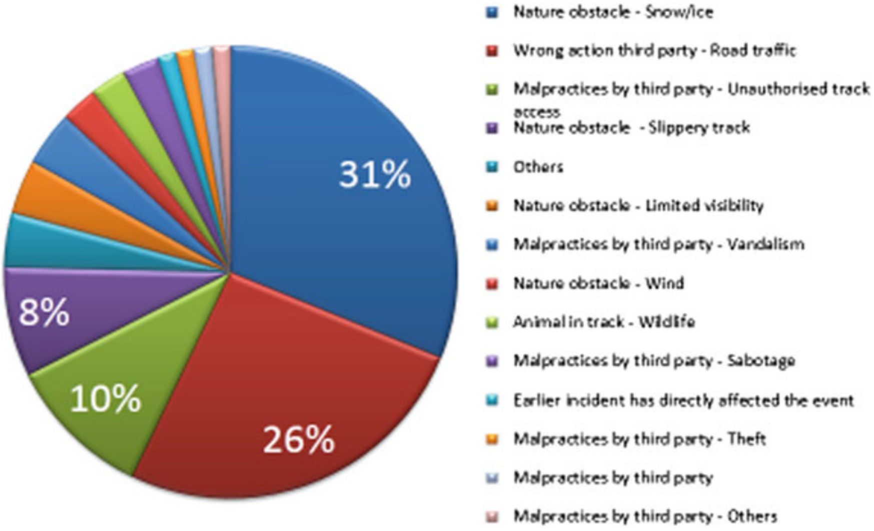


Fig. 6 Causes of external circumstances—unwanted events related to track work 2007–2012 in Sweden

3.2 *Process-Analysis Supported by HAZOP Guidewords*

In this section, the human failures classified according to the HAZOP guidewords are ranked by the frequency of investigations where they occur and related to the

Table 2 Ranking of the occurrence of human failures in investigations related to the generic maintenance process

Type of human failure	Process phase	Frequency (%) of investigations
I4	Maintenance execution, feedback	34
R4	Maintenance execution, feedback	34
A9	Maintenance execution	31
P2	Maintenance support planning	19
I4 and R4	Maintenance execution, feedback	15
C2	Maintenance execution	11
V1	Maintenance execution	7

phases of the maintenance process (see Table 2). The majority of human failures in maintenance were related to information deficiencies, i.e. information communication errors (I4, 34 %) or information retrieval errors (R4, 34 %). In 15 % of the investigations, both I4 and R4 could be identified, i.e. a combination of incorrect information and unclear communication contributed to the accidents. These information deficiencies were all located in the maintenance execution phase or in the interface between execution and other phases of the maintenance process, see Fig. 1.

The next largest group of human failures consisted of action errors (A9, 31 %). These action errors, omitted operations, are all located in the execution phase of the maintenance process.

Thereafter, the groups are in descending order: planning errors (P2, 19 %), which are located in the process phase of maintenance support planning, checking errors (C2, 11 %) located in the maintenance execution phase, or violations (V1, 7 %) located in the maintenance execution phase. The checking errors are, in addition to maintenance execution, also connected to the process phases of maintenance support planning through feedback. The violations are all located in the maintenance execution phase, but are also related to feedback to the maintenance support planning phase.

4 Discussion and Conclusions

In this paper we have applied a process model of maintenance in combination with MTO-analysis and HAZOP guidewords, in order to analyse human failures that are maintenance-related. The approach is based on a generic process view of maintenance that is intended to support continuous improvement of and continuous risk reduction in maintenance activities.

By applying a process view of maintenance and the guidewords, it is possible to facilitate the identification of human failures. This identification supports the management of both requirements and risks, which should contribute to business prosperity through continuous improvement and risk reduction.

The proposed process approach for continuous risk reduction stresses that, even though the incidents and accidents may be manifested in maintenance execution, the underlying hazards may actually be located in other process phases.

Further on, independent of what kind of philosophies, theories, and technologies that are applied within an organisation, maintenance sooner or later comes down to maintenance execution, which still requires human intervention. Hence, the effectiveness of all activities in the other maintenance phases should be evaluated by their impact on maintenance execution, especially considering human failures.

In addition to the HAZOP guidewords-inspired classification, the performed analysis is based on a generic process model. The process model provides an understanding of 'where' different causes contributing to human failures are located. Hence, the combination of the guidewords and the process approach provides information about both 'what' and 'where' aspects of causes contributing to human failures. This combination gives a preliminary understanding about the network of contributory factors in human failures during maintenance execution, i.e. 'how' the causes are interlinked with each other and together contribute to human failures. Hence, the applied analysis supports the proposition that human failures are consequences of a network of actions and conditions which involve people, teams, tasks, workplace, and organisational factors, rather than single causes of accidents. This is also achieved to some degree by system approaches, such as the MTO-analysis, but the process approach ads some further dimensions, e.g. stakeholders, deliverables, continuous improvement and their interrelationships.

The study also indicate that there are many unreported incidents and risks related to maintenance execution compared to reported accidents (703 unwanted events of which 262 accidents, 358 incidents, and 83 risks), cf. the safety pyramid as described by [36].

It may also be interesting to notice that it was not possible to identify any cause for human failure directly related to the phase of maintenance management. However, it is this phase that lay the very foundation of the other maintenance phases through financial, regulatory, and cultural governance and control. One reason for this absence might be that human failure normally is viewed in an operative perspective during maintenance execution, which seemingly makes contributory causes in maintenance management very distant. However, as discussed above, successful maintenance management must be very influential on all other maintenance process phases. The importance of committed leadership is also stressed in all major management literature. In addition, human failure should also be considered in every single phase of the maintenance process. This also points out a potential to use the proposed process approach in the investigation of unwanted events to identify further causes.

Through an application of the process approach, the performed analysis can also be seen as part of the maintenance process itself, i.e. the phase of maintenance assessment. Hence, in analogy with the process approach, the result of the analysis (or maintenance assessment) also indicates scope for maintenance improvement. These improvements address aspects of different phases of the maintenance process, as described above.

In summary, the process perspective ads some valuable dimensions to the understanding of human failures within maintenance. This is due to the general strengths of process mapping and improvements, e.g. where the focus is on activities that contribute to the achievement of deliverables to stakeholders. Hence, the process methodology can act as a facilitator to combine other supporting methodologies and tools to pinpoint improvement possibilities, such as MTO-analysis and HAZOP guidewords. Both the process and MTO-analysis approaches, facilitate a further exploration of unwanted events that occur during execution, trying to find the blunt end of the causal chain and avoid to stop at the sharp end.

Acknowledgments The financial and intellectual support received from Trafikverket (Swedish transport administration) is gratefully acknowledged.

References

1. Chambell JD, Jardine AKS (2001) Maintenance excellence: optimizing equipment life-cycle decisions. Marcel Dekker, New York
2. Liyange JP, Kumar U (2003) Towards a value-based view on operations and maintenance performance management. J Qual Maintenance Eng 9(4):333–350
3. Söderholm P, Holmgren M, Klefsjö B (2007) A process view of maintenance and its stakeholders. J Qual Maintenance Eng 13(1):19–32
4. Holmgren M, Söderholm P (2008) A process approach to maintenance-related hazard identification. Int J COMADEM 11(1):36–46
5. Akersten PA (2002) Maintenance management should be based on core values, methodologies and tools. In: Proceedings of international conference of maintenance societies, ICOMS
6. Akersten PA, Klefsjö B (2003) Total dependability management. In: Pham H (ed) Handbook of reliability engineering. Springer, London
7. HSE (2002) Train accident at Ladbroke Grove, Paddington junction—second HSE interim report. http://orr.gov.uk/__data/assets/pdf_file/0020/5663/incident-ladbrokegrove-lgri2.pdf. Accessed 2 Mar 2015
8. RAIB (2008) Track worker fatality at Ruscombe Junction 29 April 2007. Rail accident report. Rail Accident Investigation Branch, Department for Transport
9. RAIB (2009) Trackworker stuck by train, Stevenage, 7 December 2008. Rail accident report. Rail Accident Investigation Branch, Department for Transport
10. TRV (2010) 2010-09-12, Kimstad, tåg 505 kör på traktorgrävare med spårföljarhjul. Utredningsrapport TRV. Trafikverket, Borlänge
11. SHK (2912) Olycka mellan tåg 505 och en spårgående grävlastare på Kimstad driftplats, Östergötlands län den 12 september 2010. Rapport RJ 2012:03. Swedish Accident Investigation Authority, Stockholm
12. Holmgren M (2005) Maintenance–related losses at the Swedish rail. J Qual Maintenance Eng 11(1):5–18
13. Söderholm P (2013) Taking possession of the track and securing the work site faster without compromising on safety. In: Proceedings of: rail infrastructure: access planning & work window productivity
14. Dekker S (2004) Ten questions about human error. Erlbaum, Mahwah
15. Reason J (1997) Managing the risks of organizational accidents. Ashgate, Brookfield
16. Reason J, Hobbs A (2003) Managing maintenance error: a practical guide. Ashgate, Aldershot

17. Dekker S (2002) The field guide to human error investigations. Ashgate, Cornwall
18. Hollnagel E (2004) Barriers and accident prevention. Ashgate, Aldershot
19. Holmgren M (2006) Maintenance related incidents and accidents—aspects of hazard identification. Luleå University of Technology, Luleå
20. ISO 31000 (2009) Risk management—principles and guidelines. International Organization for Standardization, Geneva, Switzerland
21. ISO 22320 (2011) Societal security—emergency management—requirements for incident response. International Organization for Standardization, Geneva, Switzerland
22. ISO/IEC 27035 (2011) Information technology—security techniques—information security incident management. International Organization for Standardization, International Electrotechnical Commission, Geneva, Switzerland
23. ISO 9000 (2005) Quality management systems—fundamentals and vocabulary. International Organization for Standardization, Geneva, Switzerland
24. ISO 55000 (2014) Asset management—overview, principles and terminology. International Organization for Standardization, Geneva, Switzerland
25. IEC 60300-3-14 (2004) Dependability management—Part 3-14: application guide—maintenance and maintenance support. International Electrotechnical Commission, Geneva, Switzerland
26. COSO ERM (2015) Enterprise risk management—integrated framework. http://www.coso.org/documents/coso_erm_executivesummary.pdf. Accessed 5 Feb 2015
27. IEC 60300-3-11 (2009) Dependability management—Part 3-11: application guide—reliability centred maintenance. International Electrotechnical Commission, Geneva, Switzerland
28. IEC 60812 (2006) Analysis techniques for system reliability—procedure for failure mode and effect analysis (FMEA). International Electrotechnical Commission, Geneva, Switzerland
29. Shewhart WA (1980) Economic control of quality of manufactured product. ASQC Quality Press, Milwaukee
30. Deming WE (1993) The new economics for industry, government, education. Massachusetts Institute of Technology, Center for Advanced Engineering Study, Cambridge
31. Reason J (1990) Human error. Cambridge University Press, Cambridge
32. Reason J (1997) Managing the risks of organizational accidents. Ashgate, Aldershot
33. Rollhangen C (1995) MTO—En introduktion, sambandet människa, teknik och organisation. Studentlitteratur, Lund
34. Vinnem J-E (2001) Offshore risk assessment: principles, modelling and applications of QRA studies. Springer, London
35. HSE (2015) Core topic 3: identifying human failures. http://www.hse.gov.uk/humanfactors/topics/core3.pdf. Accessed 16 Jan 2015
36. Heinrich HW, Petersen D, Roos N (1980) Industrial accident prevention: a safety management approach. McGraw-Hill, New York

Application of Game Theory to Railway Decision Making

N. Jack, D.N.P. Murthy and U. Kumar

Abstract Over the last few decades various forms of railway privatization have taken place in many different countries. As a result, it is now common for several parties to be involved in the ownership, operation and maintenance of railway system assets. The decisions made by each impact on all others. Game theory (GT) provides the framework to obtain the optimal decisions taking into account the various interactions. This paper gives a brief introduction to GT and its application to railway decision-making.

Keywords Railway systems · Privatization · Decision problems · Maintenance · Game theory

1 Introduction

In most countries, railway assets were state-owned and operated and maintained by government agencies. Nowadays, various forms of railway privatization have taken place with organisational structures varying among different countries. As a result, it is now common for several parties to be involved in the ownership, operation and maintenance of railway system assets. Each is an independent unit with its own

N. Jack (✉)
Springfield, Cupar, Fife KY15 5SA, Scotland
e-mail: natjack51@yahoo.co.uk

D.N.P. Murthy
School of Mechanical and Mining Engineering,
The University of Queensland, St Lucia Q4067, Australia
e-mail: p.murthy@uq.edu.au

U. Kumar
Division of Operations and Maintenance Engineering,
Lulea University of Technology, Lulea 971 87, Sweden
e-mail: Uday.Kumar@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_28

goals and objectives and decisions made impact on the other parties in the system. Game theory (GT) provides the framework for each party to make their optimal decisions taking into account the various interactions.

This paper provides the conceptual basis for looking at the important central issues involved in railway systems, the appropriate framework needed to study the decision problems from the perspectives of the different parties and details on how GT can be used for optimal decision making. Section 2 gives a brief description of the railway privatization process that has taken place in Sweden and the UK. Section 3 deals with the decision problems faced by asset owners, operators and maintenance contractors in privatized railway systems. A brief introduction to GT is given in Sect. 4. The maintenance of railway infrastructure and rolling stock is described in Sect. 5. In Sect. 6, the game-theoretic approach to railway decision making is outlined together with a simple GT model formulation. Conclusions and topics for further research are given in Sect. 7.

2 Railway Privatization

Railways are complex systems comprised of infrastructure (track, sleepers and track bed, bridges, tunnels, stations, level crossings and signalling systems) and rolling stock (locomotives, wagons and carriages used to transport passengers and freight). Traditionally, they have been state-owned and operated and maintained by government agencies but nowadays, in many countries railways have been subject to various forms of privatization. This process has led to several parties (see Fig. 1) being involved in the ownership, operation and maintenance of the infrastructure and rolling stock [3]. Examples are track owners, track operators, rolling stock owners, rolling stock operators and maintenance contractors. The operators of the track and rolling stock may either be the owners of these assets or they may lease them from the owners. The operators are responsible for conducting essential maintenance and often the maintenance tasks are outsourced to external maintenance contractors.

The government of the country in which a railway system is based makes policy decisions with respect to levels of service (train routes, frequencies, etc.) and regulators are appointed mainly to ensure customer safety where the customers are the general public for passenger services and industries (e.g. mining) for freight. Finally, in the figure, the party referred to as ‘others’ includes manufacturers of rolling stock and infrastructure builders (of tracks, stations, bridges, tunnels, etc.).

Swedish Railway System

In Sweden, prior to 1988, the government-owned Swedish State Railways (SJ) controlled all aspects of railway services. The railway privatization process then began and has been slow-moving and incremental [1]. Today, most of the infrastructure is still government-owned and managed by the Swedish Transport Administration (Trafikverket).

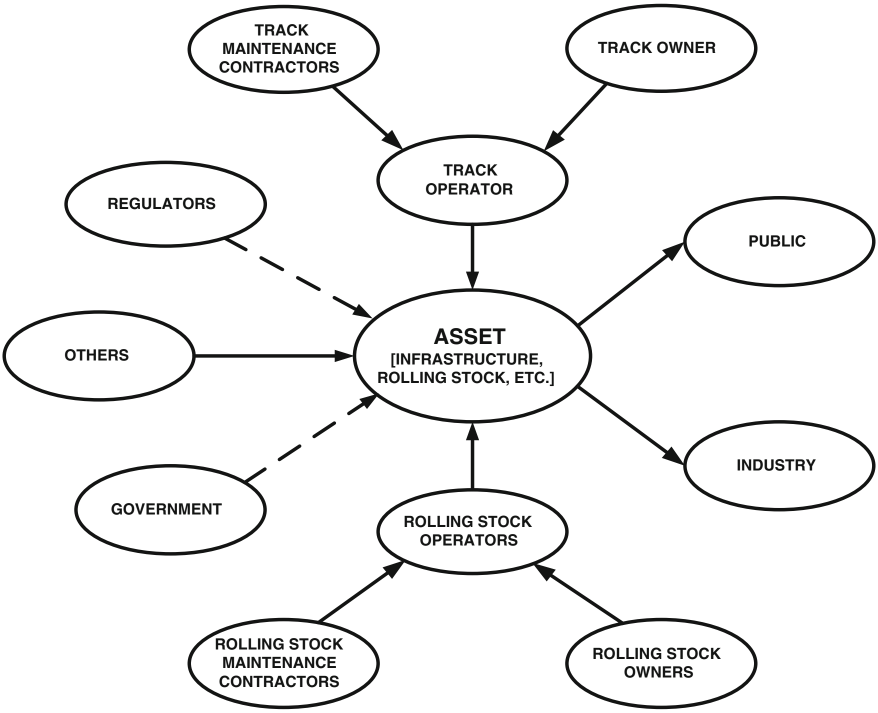

TRACK MAINTENANCE CONTRACTORS
TRACK OWNER
TRACK OPERATOR
REGULATORS
PUBLIC
OTHERS
ASSET [INFRASTRUCTURE, ROLLING STOCK, ETC.]
INDUSTRY
GOVERNMENT
ROLLING STOCK OPERATORS
ROLLING STOCK MAINTENANCE CONTRACTORS
ROLLING STOCK OWNERS

Fig. 1 The key parties involved a railway system

SJ has now been broken up into several specialized companies, many of which have been privatized whilst others are still government-owned. Presently, there are a total of 40 rolling stock companies. One of these, SJ AB is the major passenger train operator running its own services and also operating under contract to both regional and national transport authorities. SJ AB owns all its passenger trains. Inter-regional passenger services were opened up to competition in December 2011. The Swedish freight services market is now completely deregulated apart from the government-owned Green Cargo AB (previously SJ's freight division) which currently has a market share of approximately 40 %.

The Swedish Transport Agency (Transportstyrelsen) is the regulatory body responsible for the Swedish railway system. It formulates regulations, examines and grants permits, deals with safety issues and oversees the behaviour of the participating organizations.

UK Railway System

In the UK, the railway privatization process took place during the relatively short period 1994–1997 [1]. Prior to 1994, the UK railway system was operated by British Rail (BR). In 1994, the government owned company Railtrack took over the ownership and responsibility for maintaining the infrastructure. The infrastructure

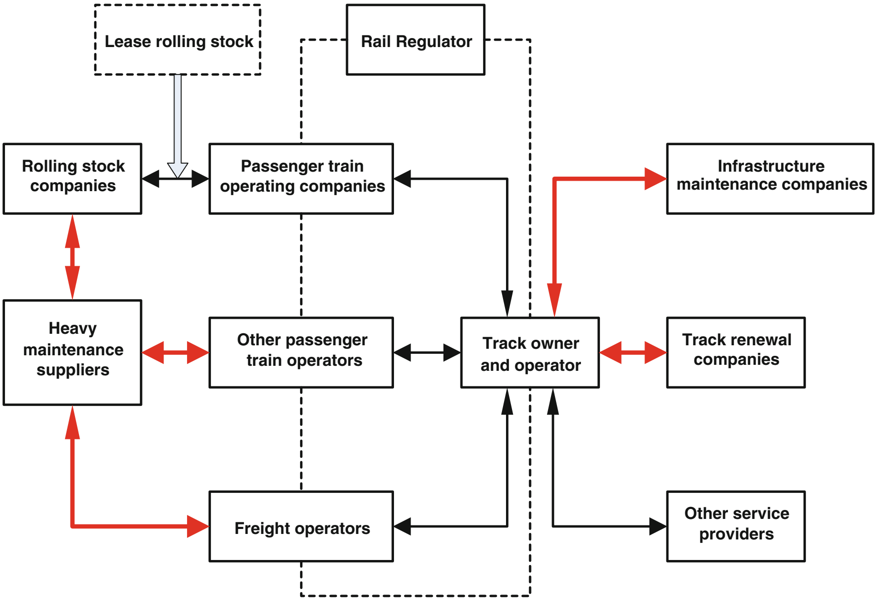


Fig. 2 Privatization of UK railways

unit and its support departments (maintenance, services design and track renewal) were then sold off to the private sector in 1996. Railtrack went into bankruptcy in 2002 and has now been effectively replaced by the state-controlled non-profit company Network Rail.

Three rolling stock companies (ROSCO's) purchased BR's passenger rolling stock and then subsequently leased the vehicles to operators. BR's freight train operations were divided into 6 companies who were then sold to the private sector. In contrast, passenger train operations were franchised to 25 private sector train operating companies (TOC's) using the newly created Office of Passenger Rail Franchising (OPRAF). This franchising body was replaced by the Strategic Rail Authority (SRA) in 2001 and then the Department of Transport (DOT) in 2005.

The UK government also set up the Office of Rail Regulation (ORR) as the safety and economic regulator for the new railway system. All the parties involved in the operation and maintenance of this system are shown in Fig. 2.

3 Railway System Decision Problems

In any privatized railway system, the owners, operators and maintenance contractors each face several decision problems. The decisions may be strategic, tactical or operational with each decision being based on data and made at different levels within an organization. The actual decision problems faced depend on the particular scenario being studied and there are several possible scenarios. The operators of the infrastructure and the rolling stock may either be the owners or they may lease the assets from the owners. We only consider the former situation since this avoids having to deal with leasing issues.

An infrastructure owner/operator has to make decisions involving the maintenance and upgrade of the different components of the infrastructure, the number of rolling stock operators to allow to use the track, the frequencies for passenger and freight transport, and whether or not to outsource the infrastructure maintenance. If the maintenance is outsourced then one or more external maintenance contractors have to be selected and contracts drafted between the different parties. Each rolling stock owner/operator has to decide about rolling stock acquisition, upgrade and replacement of the stock and also its maintenance. If maintenance outsourcing is chosen then the rolling stock owner/operator faces the same types of decision as the infrastructure owner/operator.

Contracts must also be agreed between the infrastructure owner/operator and each rolling stock owner/operator for use of the infrastructure. These contracts need to specify limits on track usage (frequency of train journeys and loads carried), acceptable conditions of the infrastructure and rolling stock, contract prices, penalties for breaching contract terms, etc. If maintenance of the infrastructure and rolling stock is outsourced, the external maintenance contractors (service agents) need to decide on the contract terms with both types of operator and on the logistics necessary to provide the maintenance services. The contracts should include details of all the maintenance activities that are to be outsourced, contract prices and penalties for not following the terms and conditions of the agreements. The contractors may also offer different contract options to the operators who then need to decide which option to choose.

Thus, each of the parties involved in privatized railway systems (the infrastructure owner/operator, the rolling stock owners/operators and the external maintenance contractors) have to evaluate a number of different options and then determine their optimal decisions, taking into account the interactions between themselves and the other parties. The solutions to these decision problems can be obtained by using appropriate models.

4 A Brief Introduction to Game Theory

GT is used to characterize optimal decision-making in problems where there are two or more decision makers. Details about GT and its various areas of application can be found in Chatterjee and Samuelson [2] and Watson [7]. The elements of a game are the players (the decision makers who participate in the game), their decision choices, and their objective functions (which depend on the outcome of the interactions that occur). The general structure of a two-player GT problem is shown in Fig. 3.

An important assumption of GT is that the players will always act rationally (choose their best decisions). In any game, an action is the decision that a player

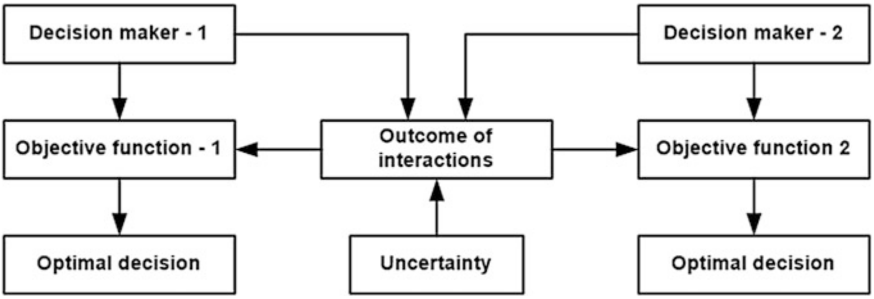


Fig. 3 Decision problem structure with two decision makers

makes at a particular point in the game whereas a strategy specifies what actions the player will take at each point in the game. A solution concept is a technique that is used to predict the outcome (equilibrium) of the game. It identifies the strategies that the players are actually likely to play in the game.

4.1 Classification of Games

GT problems may be classified in a number of different ways. The timing of actions by the players and also the number of periods during which games are played lead to different solution approaches. In some games, the players may choose their actions simultaneously, so that no player knows exactly what the others have done when they make a decision. Alternatively, in games with sequential timing, the players choose their actions in pre-determined order. These two situations are termed *Nash games* and *Stackelberg games*, respectively, and they are discussed in Sect. 4.2.

Some games take place during a single time period whereas others occur over multiple time periods and the actions taken by the players in each period affect the actions and rewards of the players in subsequent periods. These two situations are termed *static games* and *dynamic games*, respectively.

Finally, games may be either *cooperative* or *non-cooperative*. In a cooperative game the players communicate with each other to coordinate their strategies and, most importantly, make binding agreements. This type of game can be formulated as a multi-objective optimization problem. In a non-cooperative game the players may communicate, but binding agreements are not made.

4.2 Two-Player Non-cooperative Static Games

We denote the two players as P_1 and P_2. The actions available to P_1 are denoted $x \in X$ and for P_2 they are $y \in Y$. The objective functions for P_1 and P_2 are $J_1(x;y)$ and

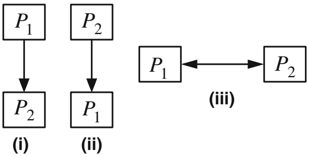


Fig. 4 Power structures for a two-person game

$J_2(y;x)$, respectively. The different power structures for a two-player non-cooperative static game are shown in Fig. 4.

Nash Games

Here, each player selects a single action without knowing the particular action chosen by their rival. This effectively means that the two players P_1 and P_2 choose their actions simultaneously and so have equal decision-making power. This power configuration corresponds to case (iii) in Fig. 4. In such a game, the players' strategies are just the single actions they choose, so the terms actions and strategies will be used interchangeably. The most well-known and widely-used solution concept for this static game is called *Nash equilibrium* (NE). A NE is a set of strategies (strategy profile) for the two players such that no player has an incentive to change their strategy unilaterally, given the strategy chosen by the other player.

More formally, the strategy profile (x^*, y^*) is a NE if

$$\begin{aligned} J_1(x^*;y^*) &\geq J_1(x;y^*), \quad \text{for all } x \in X, \text{ and} \\ J_2(y^*;x^*) &\geq J_2(y;x^*), \quad \text{for all } y \in Y. \end{aligned} \tag{1}$$

A NE may be found using *best response* functions. $P_1's$ best response $BR_1(y)$ to a given action $y \in Y$ chosen by P_2 is the value of x which maximizes $J_1(x;y)$ so

$$BR_1(y) = \underset{x \in X}{\operatorname{argmax}}\, J_1(x;y). \tag{2}$$

Similarly, $P_2's$ best response $BR_2(x)$ to a given action $x \in X$ chosen by P_1 is the value of y which maximizes $J_2(y;x)$ so

$$BR_2(x) = \underset{y \in Y}{\operatorname{argmax}}\, J_2(y;x). \tag{3}$$

For a NE, both players' actions must be best responses to each other so the NE strategy profile (x^*,y^*) is the solution of

$$x^* = BR_1(y^*) \quad \text{and} \quad y^* = BR_2(x^*). \tag{4}$$

Stackelberg Games

We assume that P_1 is the *leader* who chooses an action $x \in X$ and then P_2 (the *follower*) observes x and chooses an action $y \in Y$. This corresponds to case (i) of

Fig. 4. [Note, in case (ii) of Fig. 4, the roles are reversed, so P_2 is the leader and P_1 is the follower.]

The *backward induction* method of solution for the two-stage *Stackelberg game* depicted in case (i) is as follows.

Stage 2: Given the action x previously chosen by P_1,P_2's problem is to find the value of y that maximizes $J_2(y;x)$. The solution to this problem is the best response function

$$BR_2(x) = \arg\max_{y \in Y} J_2(y;x). \tag{5}$$

Thus, P_2 responds optimally to P_1's action.

Stage 1: P_1 anticipates what P_2 will do in stage 2, so P_1's problem in this part of the game is to

$$\max_{x \in X} J_1(x;BR_2(x)). \tag{6}$$

If x^* is the optimal solution to the optimization problem in (6) then the outcome of the game is that P_1 chooses x^* and P_2 chooses $BR_2(x^*)$.

A Stackelberg game model can be used to solve the *principal-agent problem*. This problem arises when one party (the principal) delegates tasks to another party (the agent) who then performs these tasks under a contract. The two parties have conflicting objectives and it is difficult and/or expensive for the principal to monitor the actions of the agent during the contract period. To devise the contract, the principal must include incentives/penalties to encourage the agent to behave at least partly according to the principal's interests. See Watson [7] for details of the GT model analysis of the principal-agent problem.

4.3 Other Important Issues in GT Modelling

There are various factors that make real-world decision making problems involving two (or more) players difficult to model. One such issue is the information available to each player. *Asymmetric information* implies that one player has more or superior information compared to the others. This knowledge may be about costs, revenues, etc., or actions taken during the game. *Moral hazard* is a specific type of information asymmetry that refers to effect of 'hidden actions' taken by players. This situation occurs in principal-agent problems and may be prevented by monitoring or by providing contract incentives or penalties.

A second issue is the uncertainty in the real-world that may affect the payoffs to the players plus other outcomes of a game. Each players' objective function needs to take this uncertainty into account and one way of doing this is through $E[V]$, the expected value of the monetary payoff. The disadvantages of using expected value as an objective function are that it does not take into account the variability in the possible payoff values or a player's attitude to risk. In order to capture these concepts, a utility function $U(V)$ may be utilized. This utility function is a measure of a player's preferences for different payoffs and its shape determines the player's risk attitude. In this case, each player chooses their actions in order to maximize $E[U(V)]$, the expected utility of the random payoff earned. This is known as the *principle of expected utility maximization*.

These factors often need to be taken into account in the formulation and analysis of GT models. The effect of different scenarios regarding information availability, levels of uncertainty and risk attitudes on the model solutions may then be investigated.

5 Maintenance of Railway Assets

Railway assets degrade due to age, usage (loads carried and their frequency), operational environment and other factors involved in their design and how they are operated. The interaction that takes place between the rolling stock and the track has a great influence on the degradation of both assets. Poor quality track has an adverse effect on trains, carriages and freight wagons while poor wheel quality and frequent heavy loads cause increased track deterioration.

Preventive maintenance (CM) is needed to control railway asset degradation and corrective maintenance (CM) is needed to rectify faults whenever normal operations are disrupted. In both cases, poor maintenance can also lead to accelerated asset degradation.

PM actions include ballast tamping and injection, sleeper replacement, lubrication and grinding to reduce rolling contact fatigue (RCF) initiated defects caused by wear and fatigue. Train operations are affected only for a short period for this type of maintenance. PM actions also include inspections to monitor and assess infrastructure condition. Based on the inspection results, e.g. the severity of any fault found to rail traffic and the availability of resources, a decision is made either to fix the fault immediately or plan the rectification for a later date. In contrast, CM actions may involve major maintenance work such as track reconditioning and replacement which take a much longer time (possibly running into months) to complete. Train operations are then affected significantly and the resulting CM costs are high.

The degradation of rolling stock wheels is due to profile wear and RCF. Various wayside condition monitoring technologies are used to identify these types of defect. Wayside detectors monitor bearing temperature and the forces generated by vehicles to provide alarm activation and prevent derailment. Force measurement detectors are able to detect vehicles with excessive loads and those with defective wheels that might damage the track. Wheel profile is critical to a railway vehicle's dynamic behavior, stability and the ride comfort. The rate of wear and rolling resistance are monitored automatically using laser units and cameras. When an alarm is generated by such measurements, the vehicle is identified, the damage and fault level is estimated, and maintenance is initiated [6].

If the maintenance of an asset is outsourced, some or all of the maintenance actions (PM and/or CM) are carried out by an external service agent (maintenance contractor) under a maintenance service contract. The contract specifies the details of the maintenance and the cost issues. It can be a simple or complex and may involve penalty and incentive terms. In Sweden, competitive tendering for railway maintenance contracts began in 2002 [5]. Trafikverket has divided the infrastructure in Sweden into 6 zones and outsources all the maintenance through an open tendering process where any qualified maintenance contractor can bid for the maintenance contracts. These contracts normally last for a period of 5 years with the provision for a 2 year extension given in two steps each lasting 1 year. The contract value for each zone is around 40–50 million euros. Currently, there are five private companies competing in the infrastructure maintenance market whilst eight companies compete for rolling stock maintenance contracts.

Nilsson and Nyström [4] compare railway maintenance markets and describe the design of maintenance contracts in the Netherlands, Finland and the UK.

6 Game-Theoretic Approach to Railway System Decision Making

The characterization of a railway system with one infrastructure owner/operator, one rolling stock owner/operator, one infrastructure maintenance contractor and one rolling stock maintenance contractor is shown in Fig. 5. The directed arcs indicate the interactions between the elements and the related variables.

There is an interaction between the rolling stock and the infrastructure (track) and the degradation of each asset is influenced by the interaction between them. This degradation is affected by the condition of the rolling stock and of the infrastructure and by several other factors such as load, speed of travel, etc. The infrastructure owner/operator and the rolling stock owner/operator each outsource their maintenance to a single maintenance contractor/service agent. Figure 5 also indicates the different contracts between the owner/operators and the service agents. As can be seen, several different players are involved and their decision making needs to take into account the interactions indicated in the figure.

6.1 Illustrative GT Example—A Principal-Agent Problem

Simple formulation

We consider a GT problem involving one track owner/operator (P_1), one rolling stock owner/operator for freight transport (P_2) and one track maintenance contractor (P_3).

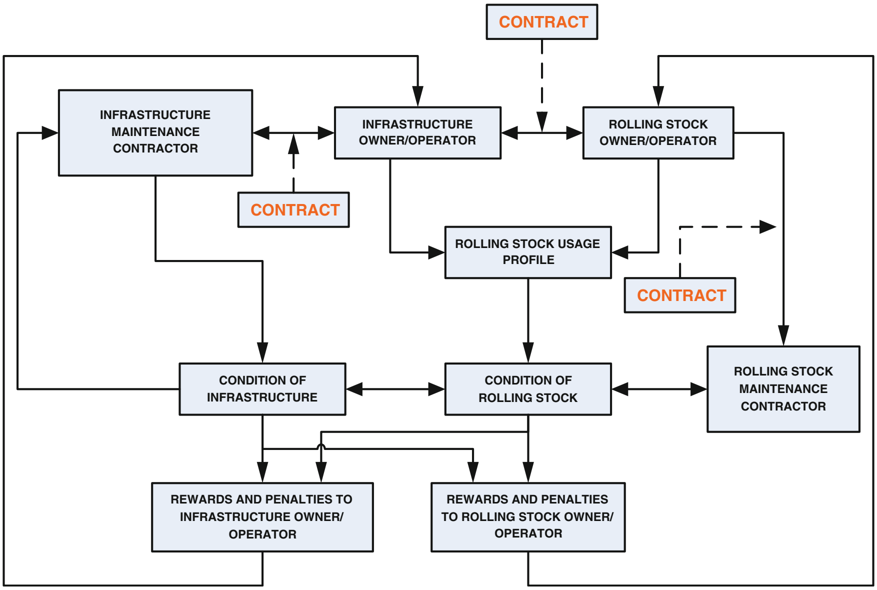

CONTRACT
INFRASTRUCTURE MAINTENANCE CONTRACTOR
INFRASTRUCTURE OWNER/OPERATOR
ROLLING STOCK OWNER/OPERATOR
CONTRACT
ROLLING STOCK USAGE PROFILE
CONTRACT
CONDITION OF INFRASTRUCTURE
CONDITION OF ROLLING STOCK
ROLLING STOCK MAINTENANCE CONTRACTOR
REWARDS AND PENALTIES TO INFRASTRUCTURE OWNER/ OPERATOR
REWARDS AND PENALTIES TO ROLLING STOCK OWNER/ OPERATOR

Fig. 5 Key elements of railway system characterization

P_1 offers one contract to P_2 which specifies the total tonnage T to be transported (contract A) and another contract to P_3 which specifies the maintenance effort M to be used (contract B). The duration of both contracts is L and their prices $\phi_A(T)$ and $\phi_B(M)$ are strictly increasing functions of T and M, respectively. P_2 and P_3 may decide to either accept or reject the contracts. If contract A is accepted then P_1 receives $\phi_A(T)$ from P_2. If contract B is accepted then P_3 receives $\phi_B(M)$ from P_1. Both contract offers and the resulting decisions are made during a single time period, so we have a static GT model formulation.

The three players are assumed to be risk neutral and each player has full information about all parameters (revenues, costs, etc.) of the game. No cheating takes place by either P_2 (transporting a tonnage $> T$) or by P_3 (using a maintenance effort $< M$), so there is no moral hazard.

The set of decision variables for P_1 is $\{T,M\}$. P_2 has the single decision variable

$$d_A = \begin{cases} 0 & \text{if contract A is rejected,} \\ 1 & \text{if contract A is accepted} \end{cases}$$

and P_3 has the single decision variable

$$d_B = \begin{cases} 0 & \text{if contract B is rejected,} \\ 1 & \text{if contract B is accepted.} \end{cases}$$

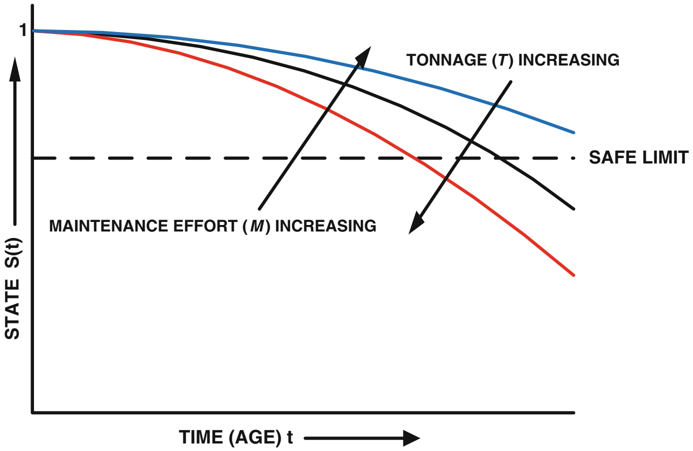


Fig. 6 Effect of tonnage and maintenance effort on track state

The objective functions for the three players are $J_{TO}(T,M;d_A,d_B)$,$J_{RSO}(d_A;T)$ and $J_{TMC}(d_B;M)$, respectively. These functions represent the profits (or expected profits) earned during the contract period and are obtained by considering revenues and costs, some of which are fixed for given T and M whilst others are subject to uncertainty.

Track degradation also needs to be modelled. There is a loss in track value during the contract period due to the tonnage transported and the maintenance effort used. This is a cost incurred by P_1 and is a function of the change in track state. The state of the track $S(t)$ at time t ($t = 0$ corresponds to the start of the contract and $S(0) = 1$) depends on the values P_1 chooses for the tonnage T and the maintenance effort M. The effect of each of these variables on $S(t)$ is indicated in Fig. 6. The model formulation for $S(t)$ may be either deterministic or stochastic.

The condition of the rolling stock also affects track degradation and this can easily be included in the modelling of track state.

Model analysis

We model the sequence of contract offers and decisions as a two-stage *Stackelberg game*. In Stage 1, the two contract offers are made simultaneously. P_1 chooses T the total tonnage to be transported under contract A and M the maintenance effort required under contract B (these variables determine the contract prices) and offers these contracts to P_2 and P_3, respectively. In Stage 2, P_2 and P_3 decide independently whether to accept or reject the two contracts (there is no strategic interaction). Rejection by either or both players ends the game and all players then make zero profits. Contract terms then need to be revised and the game repeated. Acceptance by both players implies that the two agreements come into force so the tonnage T is transported by P_2 and the maintenance effort M is carried out by P_3 for a period of duration L and profits are realised by the three players. This GT scenario is shown in Fig. 7.

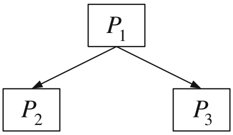


Fig. 7 Stackelberg game structure

The solution of the Stackelberg game is obtained using *backward induction*:

Stage 2: Given the values of T and M previously chosen by P_1, the best response functions for P_2 and P_3 are

$$BR_2(T) = \begin{cases} 1 & \text{if } J_{RSO}(1;T) \geq 0 \\ 0 & \text{if } J_{RSO}(1;T) < 0 \end{cases}$$

and

$$BR_3(M) = \begin{cases} 1 \text{ if } J_{TMC}(1;M) \geq 0 \\ 0 \text{ if } J_{TMC}(1;M) < 0 \end{cases},$$

respectively.

Stage 1: P_1 anticipates what P_2 and P_3 will do in Stage 2, so $P_1's$ problem in this part of the game is to find the values T^* and M^* which maximise $J_{TO}(T,M;1,1)$ subject to the two constraints $J_{RSO}(1;T) \geq 0$ and $J_{TMC}(1;M) \geq 0$.

The outcome of the game is that P_1 chooses T^* and M^*, so the contract prices are $\phi_A(T^*)$ and $\phi_B(M^*)$, and the contracts are accepted by both P_2 and P_3.

The model analysis is completed by specifying the exact expressions for the objective functions and then using the above method to obtain the optimal decisions for the three players.

Some extensions

This simple model can be extended (and so made more realistic) in many ways.

Moral hazard: This may be included so P_2 and P_3 might cheat in terms of tonnage transported and maintenance effort used during the contract period. Penalties for violation then need to be introduced by P_1 into the contract terms as extra decision variables and P_1 also needs to use monitoring to detect possible contract violation. The levels of monitoring need to be decided with increased effort resulting in higher costs to P_1 but having a greater chance of finding evidence of cheating.

Uncertainties: The revenues for P_1 and P_3 are $\phi_A(T)$ and $\phi_B(M)$, respectively, whilst the revenue for P_2 is uncertain since it depends on the demand D for freight transport (a random variable). The cost for P_1 consists of $\phi_B(M)$ plus a term that represents the loss in track value incurred during the contract period (a function of the change in track state which again may be uncertain if a stochastic formulation is used). The cost for P_2 is $\phi_A(T)$ whilst the cost for P_3 is uncertain since it depends on C the cost of the maintenance actions performed under contract B (again a random variable).

7 Conclusions and Topics for Research

The game-theoretic approach is the best method to characterize the optimal decisions of the different parties involved in privatized rail systems. In order to construct realistic GT models, proper data need to be collected from the 'real-world'. This data is essential for good model formulation and also to conduct model validation.

In this paper we have focused on the most basic GT concepts and illustrated their application through a very simple example involving three players and four decision variables. The formulation and analysis of more complex models (involving dynamic formulations, uncertainty, information asymmetry, moral hazard, etc.) is needed. The authors are currently doing further research to build such models.

Murthy and Jack [3] give an overview of the issues involved in maintenance outsourcing and leasing problems and then review the different game-theoretic models that have been proposed to help individuals/businesses choose among different options. This book provides a useful reference for both researchers and practitioners who want a better understanding of how optimal decisions are made when two or more parties are involved in the decision making process.

References

1. Alexandersson G, Hulten S (2008) The Swedish railway deregulation path. Rev Netw Econ 7:18–36
2. Chatterjee K, Samuelson WF (eds) (2001) Game theory and business applications. Kluwer Academic Publishers, Norwell
3. Murthy DNP, Jack N (2014) Extended warranties, maintenance service and lease contracts: modelling and analysis for decision making. Springer, London
4. Nilsson JE, Nyström J (2014) Mapping railway maintenance contracts—the case of Netherlands, Finland and UK, VTI report, VTI notat 27A-2014, Sweden
5. Odolinski K, Smith A (2014) Assessing the cost impact of competitive tendering in rail infrastructure maintenance services: evidence from the Swedish reforms (1999–2011). CTS working paper 2014:17
6. Palo M (2014) Condition-based maintenance for effective and efficient rolling stock capacity assurance: a study on heavy haul transport in Sweden, PhD Thesis, Division of operation and maintenance engineering, Luleå University of Technology, Luleå, Sweden
7. Watson J (2008) Strategy: an introduction to game theory, 2nd edn. W.W. Norton & Company, New York

Modelling of Maintenance Data

M.R. Karim, A. Ahmadi and D.N.P. Murthy

Abstract The modelling of maintenance data starts with the black-box approach (where the model selection is based solely on the maintenance data) and then through the grey-box approach for proper analysis (where one can gain insights to build better models). These models allow for more effective maintenance of the object. This paper deals with the grey-box approach to modelling. It discusses the process of modelling and illustrates this through a real case study.

Keywords Maintenance outsourcing · Maintenance data · Modelling · Case study · Hydraulic pumps

1 Introduction

Every engineered object is unreliable in the sense that it degrades with age and usage and ultimately fails. Preventive maintenance (PM) is used to control the degradation and reduce the likelihood of failure whilst corrective maintenance (CM) is used to restore a failed unit to the operational state. Maintenance data is the data that is collected during the maintenance of an object. It comprises reliability data (such as failure and service times), information on technical actions (such as

M.R. Karim (✉)
Department of Statistics, University of Rajshahi, Rajshahi 6205, Bangladesh
e-mail: mrezakarim@yahoo.com

A. Ahmadi
Division of Operation and Maintenance Engineering, Luleå University of Technology,
Luleå, Sweden
e-mail: alireza.ahmadi@ltu.se

D.N.P. Murthy
School of Mechanical and Mining Engineering, The University of Queensland,
St Lucia Q 4067, Australia
e-mail: p.murthy@uq.edu.au

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_29

the cause of failures), maintenance actions (such as repair or replace), economic data (such as direct and indirect costs), etc.

Maintenance data can be used to benchmark and improve the maintenance of an object. This involves proper data analysis. The two approaches that are used are the following:

(i) Qualitative: Such as Pareto charts and FMEA to identify the more frequent causes of failures.
(ii) Quantitative: Building mathematical models based on the data. Usually, one takes a black-box approach to build appropriate reliability models (such as the distribution function for failure times of the object or its components).

The reliability model built using the black-box approach often provides new insights into the degradation and maintenance of the object and this, in turn, allows one to build one or more refined models and we use the term "grey-box approach" to denote this. These models provide more information and as a result maintenance decisions based on such models lead to significant improvements (such as cost reduction) to the maintenance of the object. The paper deals with this topic and presents a real life case study to illustrate the improvements.

The outline of the paper is as follows. Section 2 deals with a brief discussion of the modelling process. Section 3 describes the case study where the engineered object is a hydraulic pump—a component of excavators used in the mining industry. It gives the data (failure and service times) that is used for building reliability models. Section 4 gives the details of the model based on the black-box approach. Section 5 looks at the grey-box approach and discusses two scenarios and the modelling of each. We conclude with some comments in Sect. 6.

2 Modelling Process

The modelling process involves a mathematical formulation which mimics the behaviour of the object being modelled. In the reliability modelling of an object the formulation used is a probability distribution function (also termed failure distribution function and which is the complement of the reliability function). Such models are useful in deciding the optimal parameters of the maintenance policy for the object.

2.1 Black-Box Approach

In the black-box approach, the selection of the model is based solely on the reliability data (failure times and service times) collected during the maintenance of an object. This approach is also referred to as *data-based* or *empirical* modelling. The modelling process involves the following steps.

Step 1: Selection of the failure distribution function $F(t;\theta)$ where θ is the set of parameters.
Step 2: Estimating the model parameters θ.
Step 3: Validating the model.

All these steps are carried out using the historical data and a variety of techniques. In Step 1 one uses various kinds of data plots—nonparametric plots such as the empirical distribution function (EDF) or parametric plots based on different distributions plots such as the Weibull probability plot (WPP). The parameter estimation in Step 2 can be done either using the method of least squares (minimising the sum of square of the errors between the model plot and data plot) or statistical methods such as the method of maximum likelihood. The validation in Step 3 can be either non-statistical (usually by comparing graphical plots) or it can be done using rigorous statistical tests such as the Anderson-Darling test. Details of these can be found in many books, such as Meeker and Escobar [7], Blischke and Murthy [2] or Blischke et al. [3] to name a few.

2.2 Grey-Box Approach

In the grey-box approach, the model based on the black-box approach is viewed from different perspectives leading to two or more scenarios. In each scenario, links are made to the real (and relevant) world of the object to gain more insight. This involves making assumptions and leads to a set of new models. The model validation often requires new additional data. If the model validity is established the new model then suggests how improvements can be made to the maintenance of the object.

3 Case Description

In open cut mines, coal and overburden are transported using excavators and dump trucks. An excavator is a complex machine consisting of several systems. The hydraulic system is one of the important systems and this is comprised of several hydraulic pumps (for linear and rotational motions), hydraulic oil filters and several hydraulic lines. Hydraulic pumps convert mechanical power into hydraulic power by delivering different flows at different load pressures at the pump output.[1]

[1] More details of the operation can be found in books on hydraulic pumps, see for example, Lambeck [6] or [5].

3.1 Pump Failures

A pump is considered to have failed if it cannot provide the required flow rate at the required pressure. Pump failure is detected by sensors and relayed to the operator. The failure is due to failure of one or more components of the pump. There can be one or more failure modes for each component and several causes leading to the failure.[2]

3.2 Pump Maintenance

The mining company used an age based policy for pump maintenance. Under this policy a pump is subjected to a replacement (PM action) after being in operation for specified period (T hours) or on failure (CM action) should it occur earlier. The pump used in the replacement may be either new or reconditioned.

Based on the condition of the pump removed (under either PM or CM) it is either scrapped or subjected to an overhaul which results in a reconditioned pump. The general accepted notion is that a reconditioned pump is as-good-as a new pump. The maintenance was outsourced to a maintenance service agent.

3.3 Data for Modelling

The mine operates 3 identical excavators on site with 2 engines per excavator and 4 hydraulic pumps (variable displacement axial piston pumps) per engine. The mine has a small maintenance department which carries out the PM and manages the outsourcing of pumps for CM actions (when a pump failure occurs) and PM actions involving the overhaul of pumps.

The data available consisted of the failure times (units that failed and removed under CM action—also called "*failure data*") and service times (units that were either still in operation or removed under PM action—also called "*censored data*") for 102 units sent to the service agent and this data are presented in Table 1. The column labelled "Type" indicates whether the data is failure data (denoted by 1) or censored data (denoted by 0). As can be seen, the data consists 45 failure data and 57 censored data. The information regarding whether an item was new or reconditioned was incomplete—15 failure data and 4 censored data are listed as new (indicated by "Yes" in the table) and 8 as reconditioned (not listed in the table). There was no information regarding the number of times a pump was reconditioned.

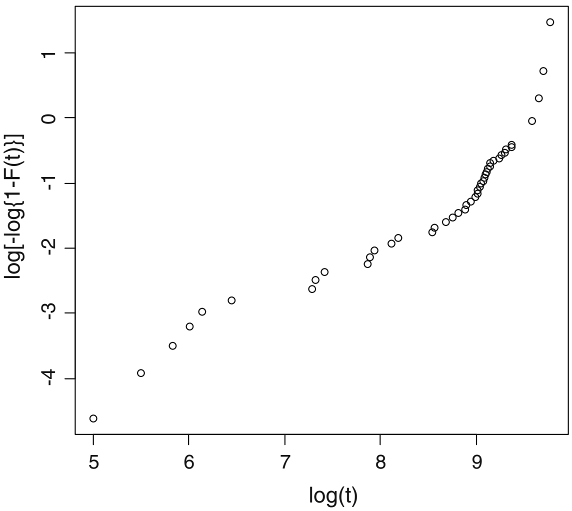
[2] For further discussion on faults, failure modes, the different physical mechanisms of degradation, etc., see [2].

Table 1 Data from customer's maintenance department

Time (h)	Type	New	Time (h)	Type	New	Time (h)	Type	New
81	0		5923	1	Yes	11923	0	
149	1		6333	1		12005	0	
245	1		6717	1	Yes	12082	0	
340	1	Yes	7207	1	Yes	12090	0	
407	1		7265	1		12136	0	Yes
461	1		7624	1	Yes	12141	0	
629	1		7625	0		12143	0	
856	0		7973	1	Yes	12163	0	
947	0		8183	1		12198	0	
1460	1		8217	1		12198	0	
1513	1		8390	1	Yes	12198	0	
1670	1	Yes	8462	1	Yes	12198	0	
1688	0		8728	1		12198	0	
2093	0		8817	1		12198	0	
2242	0		8870	1		12236	0	
2242	0		8884	0		12236	0	
2242	0		9055	1		12236	0	
2242	0		9182	1		12236	0	
2242	0		9334	1		12236	0	
2607	1		9368	1	Yes	12236	0	
2668	1		9729	1	Yes	12394	0	Yes
2806	1		9751	0		12459	0	
3132	0		10299	1		13097	0	
3132	0		10389	0		13497	0	
3132	0		10413	0		13497	0	
3132	0		10557	1		13497	0	
3333	1	Yes	10944	1		13497	0	
3569	1		10970	1		13497	0	
3837	0		11647	0	Yes	13497	0	
3837	0		11678	1		13497	0	
4150	0		11686	1	Yes	14407	1	Yes
5123	1		11798	0		15536	1	
5258	1		11869	0	Yes	16289	1	Yes
5662	0		11869	0		17517	1	

Engine number was recorded but not the location of the pump in relation to the engine. The maintenance department did not have any information regarding the failure mode. This might be due to the maintenance service contract not dealing with this issue. The service agent refused to provide this data and it is not sure whether this kind of data was collected or not.

Fig. 1 WPP plot of EDF



4 Black-Box Models

The empirical WPP obtained using the data is shown in Fig. 1 and as can be seen the plot is not a straight line. This implies that the two-parameter Weibull distribution is not appropriate to model the failure distribution.[3]

However, the shape of the plot in Fig. 1 indicates that a Weibull mixture distribution might be an appropriate model.[4] The cumulative distribution functions (CDFs) for two- and three-fold Weibull mixtures, respectively, are given by

$$G_2(t) = p_1F_1(t) + p_2F_2(t) \tag{1}$$

with $0 \leq p_1, p_2 \leq 1$ and $p_1 + p_2 = 1$, and

$$G_3(t) = p_1F_1(t) + p_2F_2(t) + p_3F_3(t) \tag{2}$$

with $0 \leq p_1, p_2, p_3 \leq 1$ and $p_1 + p_2 + p_3 = 1$, and $F_i(t),\ i = 1, 2, 3$, are the CDFs of Weibull distributions with

$$F_i(t; \alpha_i, \beta_i) = 1 - e^{-(t/\alpha_i)^{\beta_i}}, \quad t \geq 0 \tag{3}$$

where β_i is shape parameter and α_i is scale parameter, i = 1, 2, 3.

[3] Similar plots (produced using Minitab) for ten other distributions also indicate that the shapes are not straight lines.

[4] [9] deal with various models derived from the two-parameter Weibull distribution to model failure times and give the WPP plots for several distributions.

We also consider a single Weibull distribution given by

$$G_1(t) = F_1(t; \alpha_1, \beta_1) = 1 - e^{-(t/\alpha_1)^{\beta_1}}, \quad t \geq 0. \tag{4}$$

4.1 Parameter Estimation

The parameters of the models were estimated using the method of maximum likelihood. The Expectation-Maximization (EM) algorithm was applied to find the maximum likelihood estimates (MLEs) of the parameters. Table 2 shows the MLEs of the model parameters.

Comment: The mean for a Weibull distribution with shape parameter β and scale parameter α is given by $\alpha\Gamma(1 + 1/\beta)$ where $\Gamma(\cdot)$ is the gamma function [1]. For the 3-fold Weibull mixture distribution, the mean for $F_3(t)$ > mean for $F_2(t)$ > mean for $F_1(t)$. Also note that the shape parameter for $F_1(t)$ is close to one.

4.2 Model Validation

The statistical approach provides a more rigorous method for model selection and validation. Various statistics (such as Anderson–Darling (AD), Kolmogorov-Smirnov (K-S)) as well as criteria (such as the Akaike information criterion (AIC), root-mean-square error (RMSE) and log-likelihood) form the basis for model selection and validation. The adjusted AD, K-S, AIC and RMSE for the three distributions are given in Table 3.

The estimates of the adjusted AD, AIC and RMSE in Table 3 suggest that the 3-fold Weibull mixture distribution is the best distribution for the data among the three competing distributions.

Table 2 MLEs of model parameters

Model	MLEs of parameters
Single Weibull	$\{\hat{\beta}_1, \hat{\alpha}_1\} = \{1.2222, 16639.7\}$
2-fold Weibull	$\left\{\hat{\beta}_1, \hat{\alpha}_1, \hat{\beta}_2, \hat{\alpha}_2, \hat{p}_1, \hat{p}_2\right\} = \{1.0707, 1489.5746, 2.7362, 15008.6136, 0.1198, 0.8802\}$
3-fold Weibull	$\left\{\hat{\beta}_1, \hat{\alpha}_1, \hat{\beta}_2, \hat{\alpha}_2, \hat{\beta}_3, \hat{\alpha}_3, \hat{p}_1, \hat{p}_2, \hat{p}_3\right\} = \{1.019, 2364.207, 5.576, 9481.855, 16.643, 16535.503, 0.166, 0.322, 0.512\}$

Table 3 Estimated adjusted AD, K-S, AIC and RMSE for the three models

Model	AD (adj)	K-S	AIC	RMSE	Log-likelihood
Single Weibull	3.985	0.332	976.384	0.078	−486.192
2-fold Weibull	1.371	0.178	970.574	0.035	−480.287
3-fold Weibull	0.627	0.107	965.594	0.025	−474.797

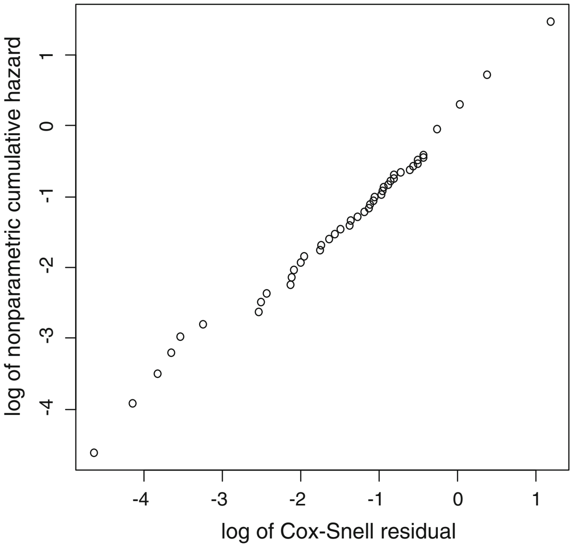


Fig. 2 Cox-Snell residuals plot for 3-fold Weibull mixture model

We use the Cox-Snell residual plots to validate the model.[5] The plot is given in Fig. 2, for the model involving the 3-fold Weibull mixture distribution. As can be seen, the residuals follow a straight line with unit slope, indicating a good fit of the three-fold Weibull mixture distribution for the data set.

We apply the Kolmogorov-Smirnov test as a goodness-of-fit test for the three-fold Weibull model. With $n = 102$, the critical value of the Kolmogorov-Smirnov one-sample test at the 5 % level of significance is $1.36/\sqrt{102} = 0.135$ [10]. Since for the 3-fold Weibull mixture distribution, the observed value of the K-S test statistic, 0.107 (given in Table 3), is less than the critical value, we cannot reject the null hypothesis, H_0, that the observed data are from a population specified by the 3-fold Weibull mixture distribution. For the other two (one- and two-fold) models the test indicates that they should both be rejected.

[5] For more on Cox Snell residuals, see, Cox and Snell [4] and [7].

5 Grey-Box Models

As discussed in the previous section the black-box approach to modelling the failure distribution indicated that a three-fold Weibull mixture (given by $G_3(t)$) is the most appropriate. This suggests that the data has come from three sub-populations. Each of these sub-populations can be interpreted in terms of the characterisation of the real world relevant to the problem. We indicate the characterisation in terms of the assumptions made and this leads to model building based on the grey-box approach—combining data with the new insights obtained from the black-box approach to modelling. We look at the following two scenarios.

5.1 Scenario 1

This is based on the following assumptions:

1. All new pumps are statistically identical
2. Some of the items replaced during PM and CM action (or service exchange) are scrapped (as they are deemed to be repairable) and others reconditioned.
3. All reconditioned pumps are also statistically identical
4. The reliability characteristics of a new pump are different from that of a reconditioned pump
5. A pump used during service exchange can be either correctly or incorrectly installed.

5.1.1 Model Formulation

We use the following notation:

q:	Probability that the item is scrapped and replaced by a new one under service exchange
p:	Probability that the item used in service exchange is installed correctly
$F_N(t)$:	Failure distribution of new item installed correctly
$F_R(t)$:	Failure distribution of reconditioned item installed correctly
$F_I(t)$:	Failure distribution of incorrectly installed item (new or reconditioned)

Comment: (1-*q*) is the probability that the item is not scrapped and reconditioned under service exchange) and (1-*p*) is the probability that the item under service exchange is not installed correctly.

As a result, the probabilities of the different outcomes after a service exchange are as indicated in Table 4.

Table 4 Probabilities of different outcomes

			Installation	
			Correct	Incorrect
			p	$(1-p)$
Scrap/repair	Scrap (new)	q	qp	$q(1-p)$
	Not scrap (recondition)	$1-q$	$(1-q)p$	$(1-q)(1-p)$

It is easily seen (using the conditional approach) that the time to failure of an item used in service exchange is given by a distribution function

$$\tilde{G}_1(t) = (1-p)F_I(t) + (1-q)pF_R(t) + qpF_N(t) \tag{5}$$

We assume that all the three distributions in the right side of (2) are two-parameter Weibull distributions. Then (5) is identical to (2) with the following equivalence among the parameters and the various distribution functions:

$$p_1 = (1-p),\ p_2 = (1-q)p \,\text{and}\, p_3 = qp \tag{6}$$

$$F_1(t) = F_I(t),\ F_2(t) = F_R(t) \,\text{and}\, F_3(t) = F_N(t) \tag{7}$$

We can obtain estimates of p and q from (6) and the estimates of the parameters for the three-fold Weibull given in Table 2. This yields $\hat{p} = 1 - \hat{p}_1 = 1 - 0.166 = 0.834$ and $\hat{q} = 0.614$.

Also note that the MTTF (mean time to failure) for a new item installed correctly > MTTF for a reconditioned item installed correctly > MTTF for an item (new or reconditioned) installed incorrectly.

5.1.2 Model Validation

The validation of a model can be achieved by checking if the assumptions made and the implications of model are valid or not. This requires additional data and tests. The tests can be based on either an intuitive approach or a rigorous statistical test.

- The model implies that the probability of an item (under service exchange) being scrapped is $\hat{q} = 0.614$. In other words, around 60 % of the items removed under service exchange are scrapped. If the service provider had collected information regarding number of items scrapped as a fraction of the items removed under service exchange and if it differed significantly from the estimate then the model is not a valid model and needs to be rejected.
- The model implies that probability of an item being installed incorrectly (during service exchange) is $(1-\hat{p}) = 0.166$. Some of the causes (such as

misalignment) can be viewed as an incorrect installation. By properly defining the other causes that result in incorrect installation this would allow one to see if the estimates obtained from the new data are close to the estimates from the model.

Comment: If the data collected included the number of failed items scrapped and failure modes then one can obtain estimates of p and q from this data. These could then be compared with the estimates $\hat{p} = 0.834$ and $\hat{q} = 0.614$. If they are in reasonable agreement then it would provide a validation whether the assumptions of Scenario 1 are correct or not.

5.2 Scenario 2

This scenario is based on the following assumptions:

1. The pumps are not all identical and can be divided into two groups—GR_1 (more reliable) and GR_2 (less reliable). This could happen, for a variety of reasons two of which are as follows:
   - New pumps used in the maintenance are bought from two manufacturers. The pumps from the first manufacturer belong to GR_1 and those from the second belong to GR_2.
   - The field reliability of a pump depends on its location in the excavator. We assume that pumps in some specific locations belong to GR_2 and the others to GR_1.
2. New and used pumps are statistically similar (reconditioning is perfect—this is commonly accepted in the mining industry but never proven rigorously).
3. Pumps used in GR_1 have failure times from a distribution function different from that used in GR_2
4. A pump used during service exchange (under PM or CM action) is either correctly or incorrectly installed.

We use the following notation:

q: Probability that the pump was used in an engine is from group GR_1

p: Probability that the item used in service exchange is installed correctly

$F_{GR1}(t)$: Failure distribution of pump used in an engine from GR_1 and installed correctly

$F_{GR2}(t)$: Failure distribution of pump used in an engine from GR_2 and installed correctly

$F_I(t)$: Failure distribution of incorrectly installed item (from GR_1 or GR_2)

5.2.1 Model Formulation

Then, following the approach used in Scenario 1, we have

$$\tilde{G}_2(t) = (1-p)F_I(t) + qpF_{GR1}(t) + (1-q)pF_{GR2}(t) \tag{8}$$

We can obtain estimates of p and q in a manner similar to Scenario 1.

5.2.2 Model Validation

As with Scenario 1, the model validation requires checking the validity of the assumptions made and this requires additional data and tests. The ability to clearly identify which of the two plausible reasons is valid or not needs additional data. In the first case, if all the pumps are bought from the same supplier then one needs to look if changes to manufacturing led to two groups—for example, pump reliability increasing (with design improvements) or decreasing due to ineffective quality control. If not, then pumps need to be identified by the manufacturer. In the second case, one needs information regarding the location of each pump within an excavator and the excavator.

5.3 Improvements to Pump Maintenance

The advantage of using models based on the grey-box approach over the black-box approach can be seen in terms of the asymptotic expected cost per unit time for the age based maintenance policy given by

$$J(T;F(\cdot)) = \frac{F(T)C_f + R(T)C_p}{\int\limits_0^T R(t)dt} \tag{9}$$

where C_p is the average cost of a PM action, C_f is the average cost of a CM action, and $R(t) = 1 - F(t)$. T^*, the optimal T, is the value that yields a minimum for $J(T;F(\cdot))$.

The optimal T's depend on the average cost of each CM and PM. We use the following additional notations.

C_n: Sale price for new pump ($80,000)

C_r: Cost (charged by the service agent) for reconditioning a pump under CM or PM action ($60,000)

ζ: Additional cost (due to downtime, loss in revenue, etc.) resulting from CM action. We look at values of ζ = $70,000, 90,000 and 110,000

We consider the following five cases:

Case (i): A maintenance action involves replacement by a new item or a reconditioned item with probabilities $\hat{q}$ (= 0.614) and $1-\hat{q}$ (= 0.316) respectively. As a result, the average cost of a PM action is $C_p = \hat{q}C_n + (1-\hat{q})C_r$ and of a CM action is $C_f = C_p + \zeta$. The optimal T^* is obtained using (9) with $F(t) = G_3(t)$ and the optimal expected cost per unit time is given by $J(T^*; G_3(\cdot))$.

Case (ii): PM action is based on Scenario 1 with uncertainty in the installation process. In this case, the optimal PM intervals for new and reconditioned items are different. For new items it is T_1^* obtained from (9) with

$$F(t) = \tilde{F}_N(t) = vF_N(t) + (1-v)F_I(t) \tag{10}$$

with $F_N(t) = F_3(t)$, $F_I(t) = F_1(t)$ and $v = 1 - p_1 = 0.834$. For reconditioned items, it is T_2^* obtained from (9) with

$$F(t) = \tilde{F}_R(t) = vF_R(t) + (1-v)F_I(t) \tag{11}$$

with $F_R(t) = F_2(t)$, $F_I(t)$ and v is as before.

Here we treat new items different from reconditioned items and so the failure distributions are different (in both cases—it is a two-fold mixture due to imperfect installation).

For new items: $C_p = C_n$ and T_1^* is obtained using this and $C_f = C_p + \zeta$.

For reconditioned items: $C_p = C_r$ and T_2^* is obtained using this and $C_f = C_p + \zeta$.

Since both new and reconditioned items are used, the optimal expected cost per unit time is given by

$$J(T_1^*, T_2^*; \varphi) = \varphi J(T_1^*; \tilde{F}_N(\cdot)) + (1-\varphi)J(T_2^*; \tilde{F}_R(\cdot)) \tag{12}$$

with $\varphi = \hat{q}$.

Case (iii): PM action is based on Scenario 1 and every item (under CM or PM action) is installed correctly. This would require proper training of the technicians to avoid causes leading to incorrect installation. In this case, the optimal T^* for new and reconditioned items, denoted by T_1^* and T_2^*, are obtained using (10) and (11) respectively with $v = 1$ (perfect installation).

The optimal expected cost per unit time is given by

$$J(T_1^*, T_2^*; \varphi) = \varphi J(T_1^*; F_N(\cdot)) + (1-\varphi)J(T_2^*; F_R(\cdot)) \tag{13}$$

with $\varphi = \hat{q}$.

Case (iv): PM action is based on Scenario 2 with uncertainty in the installation process and the group membership of each item is known. In this case, the optimal PM intervals for items from the two groups are different. For items from GR_1 it is T_1^* which is obtained from (9) with

$$F(t) = \tilde{F}_{GR1}(t) = \tilde{v}F_{GR1}(t) + (1 - \tilde{v})F_I(t) \tag{14}$$

with $F_{GR1}(t) = F_3(t)$, $F_I(t) = F_1(t)$ and $\tilde{v} = 1 - p_1 = 0.834$. For items from GR_2 it is T_2^* which is obtained from (9) with

$$F(t) = \tilde{F}_{GR2}(t) = \tilde{v}F_{GR2}(t) + (1 - \tilde{v})F_I(t) \tag{15}$$

where $F_{GR2}(t) = F_2(t)$ and $\tilde{v}$ is as before.

Since there are two manufacturers (one more reliable and the other less reliable), items from GR_1 and GR_2 have different failure distributions (in both cases—it is a two-fold mixture due to imperfect installation).

Here we assume reconditioned items are as-good-as-new. Since some items get scrapped, the average cost of a PM action is given by $C_p = \hat{q}C_n + (1 - \hat{q})C_r$

The optimal expected cost per unit time is given by

$$J(T_1^*, T_2^*; \varphi) = \varphi J(T_1^*; \tilde{F}_{GR1}(\cdot)) + (1 - \varphi)J(T_2^*; \tilde{F}_{GR2}(\cdot)) \tag{16}$$

with $\varphi = \hat{q}$.

One would expect the sale purchase price for the new items to be slightly higher. Thus, C_n for GR_1 is \$85,000 and for GR_2 its is \$80,000. As before, $C_f = C_p + \zeta$.

Case (v): PM action is based on Scenario 2 and every item (under CM or PM action) is installed correctly as discussed in Case (iii). In this case, the optimal T^* for GR_1 items and GR_2 items, denoted by T_1^* and T_2^*, are obtained from (9) using (14) and (15) respectively with $v = 1$ (perfect installation).

The optimal expected cost per unit time is given by

$$J(T_1^*, T_2^*; \varphi) = \varphi J(T_1^*; F_{GR1}(\cdot)) + (1 - \varphi)J(T_2^*; F_{GR2}(\cdot)) \tag{17}$$

with $\varphi = \hat{q}$.

In each case, the optimal $T's$ depend on ζ. The optimal $T's$ for Cases (i)–(v) are given in Table 5. As can be seen, the optimal $T's$ decrease with ζ increasing as to be expected.

Table 6 gives the optimal expected cost per year for the five cases. The percentage reduction in the cost for Case (j), j = ii, iii, iv, v, is given by the following expression:

$$100 \times \{\text{Cost for Case(i)} - \text{Cost for Case(j)}\}/\text{Cost for Case(i)}.$$

Table 5 Optimal *T*'s for Cases (i)–(v)

Case	Optimal parameters		
	$\zeta = 90000$	$\zeta = 110000$	$\zeta = 130000$
(i)	$T^* = 14485$	$T^* = 14378$	$T^* = 14296$
(ii)	$T_1^* = 14188$	$T_1^* = 14045$	$T_1^* = 13932$
	$T_2^* = 7097$	$T_2^* = 6882$	$T_2^* = 6713$
(iii)	$T_1^* = 13920$	$T_1^* = 13753$	$T_1^* = 13615$
	$T_2^* = 7086$	$T_2^* = 6833$	$T_2^* = 6293$
(iv)	$T_1^* = 14145$	$T_1^* = 14004$	$T_1^* = 13892$
	$T_2^* = 7308$	$T_2^* = 7081$	$T_2^* = 6902$
(v)	$T_1^* = 13870$	$T_1^* = 13703$	$T_1^* = 13566$
	$T_2^* = 6957$	$T_2^* = 6708$	$T_2^* = 6508$

Table 6 Optimal normalised expected cost per year for Cases (i)–(v)

	$\zeta = 90000$		$\zeta = 110000$		$\zeta = 130000$	
Case	Optimal cost/year ($\times 10^4$)	% Reduction (%)	Optimal cost/year ($\times 10^4$)	% Reduction (%)	Optimal cost/year ($\times 10^4$)	% Reduction (%)
(i)	9.1462		9.9620		10.773	
(ii)	8.4864	7.21	8.9707	9.95	9.4371	12.40
(iii)	6.2965	31.16	6.4458	35.30	4.5483	57.78
(iv)	8.9105	2.58	8.9105	10.55	8.9105	17.29
(v)	6.8047	25.60	6.9419	30.32	7.0609	34.46

Note that the costs for Case (iii) < for Case (ii) < for Case (i) and similarly the cost for Case (v) < for Case (iv) < for Case (i) for all values of ζ as to be expected. Cases (iii) and (v) correspond to the service agent making improvements to eliminate failures due to incorrect installation and as a result a significant reduction in the annual maintenance cost.

6 Conclusions

As can be seen from the case study the models based on grey-box approach result in lower maintenance costs to the owner of the object. The use of such models requires additional data to establish the validity of the model. This implies that data collection needs to be done properly. When maintenance is outsourced then this issue needs to be addressed properly in the maintenance service contract as discussed in [8]. This requires joint partnership with proper incentives so that it leads to a win-win situation for the all parties involved. There is need for further research into this topic.

References

1. Abramowitz M, Stegun IA (1972) Handbook of mathematical functions. Dover, New York
2. Blischke WR, Murthy DNP (2000) Reliability—Modeling, prediction and optimization. Wiley, New York
3. Blischke WR, Karim MR, Murthy DNP (2011) Warranty data collection and analysis. Springer, London
4. Cox DR, Snell EJ (1968) A general definition of residuals (with discussion). J Roy Stat Soc B 30:248–275
5. Krivchenko GI (1994) Hydraulic machinery: turbines and pumps. Lewis Pub, Boca Raton
6. Lambeck RP (1983) Hydraulic pumps and motors: selection and application for hydraulic power control. Marcel Dekker, New York
7. Meeker WQ, Escobar LA (1998) Statistical methods for reliability data. Wiley, New York
8. Murthy DNP, Karim MR, Ahmadi A (2015) Data management in maintenance outsourcing. Reliability engineering and system safety 142:100–110
9. Murthy DNP, Xie M, Jiang R (2004) Weibull models. Wiley, New York
10. Siegel S, Castellan NJ (1988) Nonparametric statistics for the behavioral sciences. McGraw-Hill, New York

Scheduling of Railway Infrastructure Maintenance Tasks Using Train Free Windows

Stephen M. Famurewa, Arne Nissen and Uday Kumar

Abstract Condition based maintenance scheduling is a promising approach towards effective track possession management in railway transport. If the maintenance tasks arising from condition monitoring and inspection of railways are efficiently scheduled, high service quality and capacity would be assured. In this paper, the authors presents a short-term maintenance scheduling problem to efficiently use available train-free periods for restoration of potential failures such that availability and capacity are maximised. The formulated problem focuses on reducing the possession cost and penalty cost. It is modelled as a quadratically constrained mixed integer programming problem and solved using a branch and cut algorithm. A case study on the Swedish iron ore line is used to demonstrate the use of the scheduling approach for effective track possession management.

Keywords Inspection remarks · Maintenance · Schedule · Railway infrastructure

1 Introduction

The expansion of economic activities and the increasing mobility of people have led to higher axle loads, increased speeds and tighter train movements that leave little room for daytime maintenance. Therefore infrastructure managers (IM) are concerned with increasing the competitiveness of railway transport through capacity and service quality enhancement. For example effective track possession manage-

S.M. Famurewa (✉) · U. Kumar
Luleå University of Technology, Luleå, Sweden
e-mail: stefam@ltu.se

U. Kumar
e-mail: uday.kumar@ltu.se

A. Nissen
Trafikverket, Luleå, Sweden
e-mail: arne.nissen@trafikverket.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_30

ment for maintenance plays a key role for supporting the design capacity of existing networks without compromising safety and quality requirements.

Track possession for maintenance and renewal of railway infrastructure varies, depending on the type of work, required resources and machinery. Generally, the possession requirement for track works can be summarised by adapting the conventional maintenance overview in the railway transport context, as shown in Fig. 1. In other words, possessions for track works include possession for inspection and restoration of potential failure, possession for immobilising or functional failure, possession for large–scale or planned tasks and possession for renewal works.

Maintenance planning and scheduling are essential elements of the maintenance management process which defines the tasks to be performed, analyses them to determine the required information and resources, and identifies and assigns the needed support efficiently [1]. An overview of the general structuring of the railway infrastructure maintenance planning process, state-of-the-art view on degradation modelling and task scheduling for track possession were presented by Dekker and Budai [2].

It is a common practice among the railway IMs in Europe to request and plan long possession periods for maintenance 18–24 months in advance to ensure minimal disruption to traffic [4]. However, short possessions periods are requested within short timescales to restore potential failures reported during inspection and condition monitoring. These inspections include visual inspection or non-destructive testing such as ultrasonic inspections, eddy current check, track geometry measurement and laser inspections [5–7]. Largely, inspection and condition monitoring of railways are based on factors such as the traffic volume, line speed, geotechnical conditions, technical structure, design considerations, age and

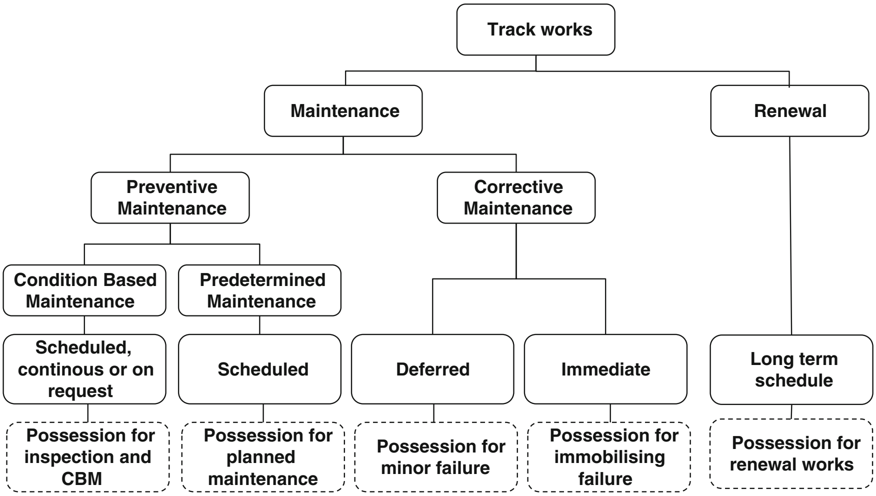

Track works
Maintenance
Renewal
Preventive Maintenance
Corrective Maintenance
Condition Based Maintenance
Predetermined Maintenance
Scheduled, continous or on request
Scheduled
Deferred
Immediate
Long term schedule
Possession for inspection and CBM
Possession for planned maintenance
Possession for minor failure
Possession for immobilising failure
Possession for renewal works

Fig. 1 Possession requirement for maintenance and renewal of railway infrastructure (adapted from [3])

other conditions related to operation. The reported potential failures or remarks are classified into different priority levels (acute, weekly, monthly, next inspection and yearly based) based on the actual condition and risk of not fulfilling the expected functional performance. This priority system is the basis for scheduling maintenance tasks with available resources such that the most urgent and important tasks are performed first.

Different studies in the past have contributed towards effective railway maintenance planning, scheduling and track possession management. Higgins [8] addressed an aspect of the possession problem for determining the best allocation of railway maintenance activities and crew to minimize train disruption. A methodology for dividing a railway network into working zones that will be taken out of service to carry out maintenance activities was presented by van den Hertog et al. [9]. Cheung et al. [10] developed a track possession assignment programme for assigning railway tracks to a given set of scheduled maintenance tasks considering defined constraints. A time-space network model was presented by Peng et al. [11] to solve the track maintenance scheduling problem by minimizing the total travel costs of the maintenance teams and the impact of maintenance projects on railroad operation.

Budai et al. [12] presented a preventive maintenance scheduling programme to combine routine tasks and projects for a link over a definite period such that the sum of possession costs and maintenance costs is minimised. Zhang et al. [13] developed a maintenance cost model and suggested an enhanced genetic algorithm approach to produce an optimal monthly schedule for maintenance works of one or more teams assuming that the deterioration of the segments are probabilistic. Forsgren et al. [14] developed a mixed integer programming model that optimises a production plan and suggests the best possible traffic flow given a fixed set of planned maintenance activities.

There is need to extend the existing scheduling models to cover maintenance tasks for potential failures and deferrable failures, which are reported in short time and not included in long term maintenance plan. To this end, this article presents an analysis of potential failure reports and an approach for the use of available train-free possession windows for maintenance.

2 Method

The method employed in this paper involves formulation of a short term maintenance scheduling model that can be used for possession management of potential failures. The track section is divided into 10 maintenance segments for logistic and operational purposes. Only the segment used for maintenance is considered occupied during a given window; thus only tasks on the same segment can be merged during a window to prevent shutting down of the entire section and to avoid too long travelling times.

2.1 Model Formulation

The objective function minimizes the total maintenance cost, which is the sum of the direct and indirect maintenance costs. The direct maintenance cost consists of two cost elements namely fixed cost per task and labour cost that depends on the estimated possession time required for each task. The indirect cost consists of three cost elements namely variable possession cost, fixed window start-up cost and penalty cost. The objective function is given by Eq. (1).

$$\begin{gathered}\min \sum_{m \in M} \sum_{w \in W} c_{mw} x_{mw} \\ M = 1, 2, \ldots .m_T, W = 1, 2, \ldots .w_T\end{gathered} \tag{1}$$

where c_{mw} is the aggregated cost of using window w on day d_w for task m with deadline on day D_m. x_{mw} is the decision variable for carrying out task m during window w. x_{mw} is a binary variable where 1 means that task m is implemented in window w and 0 means otherwise. The aggregated cost of using window w for task m is the sum of the direct maintenance cost, possession cost, window start-up cost and penalty cost. These cost elements are estimated using the formulation in Eqs. 2–5.

$$\textit{direct maintenance cost } = c_m t_m + c_{fm} \tag{2}$$

$$\textit{possession cost } = c_w t_m \tag{3}$$

$$\textit{window startup cost } = c_{fw} \mathrm{F}(\mathrm{x}_{mw}) \tag{4}$$

where $\mathrm{F}(\mathrm{x}_{mw}) = \min\left(\sum_m x_{mw}, 1\right) \quad w \in W$

$$\text{penalty cost} = c_p \max(d_w - D_m,\ 0) \tag{5}$$

In explicit terms, c_{mw} depends on the time t_m required for implementing task m, labour cost per hour c_m, fixed cost for each task c_{fm}, hourly cost for window possession c_w, fixed window start-up cost c_{fw} and the daily penalty cost c_p for exceeding the task deadline. The simplified expression given in Eq. 6 is used to estimate the aggregated cost c_{mw} for carrying out maintenance task m using window w.

$$\begin{gathered}c_{mw} = (c_m + c_w) t_m + c_{fm} + c_{fw} \mathrm{F}(\mathrm{x}_{mw}) + c_p \max(d_w - D_m, 0) \\ m \in M \text{ and } w \in W\end{gathered} \tag{6}$$

An important aspect of this model is penalty cost modelling. Penalty cost is introduced to efficiently use available windows so that critical tasks are not unduly delayed. The conventional practice during inspection is to assess the infrastructure condition and provide priority remarks on the urgency of required intervention. This can be taken as limit which when exceeded requires extra measures that decrease the capacity/performance of the concerned segment. The daily penalty cost adopted is calculated from the estimated delay consequences based on reduction of line speed from 120 to 70 km/h. In the case study, a penalty cost of 1000 € per day is imposed.

The objective function is subject to the constraints explained below:

Constraint 1: Implementation of maintenance tasks within a window should not exceed the window duration. This is presented by Eq. 7, where $\mathrm{t_m}$ is the time required to fix remark m, $\mathrm{t_w}$ is the duration of window w and $\mathrm{Trt_{mm'}}$ is the time required to travel between the locations of two task m and m′ on the same segment. An average travelling time of 10 min is used in the case study.

$$\sum_{\substack{m, m' \in M \\ m \neq `m'}} t_m \cdot x_{mw} \leq t_w - Trt_{mm'}, \quad w \in W \tag{7}$$

Constraint 2: All tasks must be completed, i.e. a task expected to take $\mathrm{t_m}$ hours should have a total sum of $\mathrm{t_m}$ hours. This constraint is defined by Eq. 8.

$$\sum_{w \in W} x_{mw} = 1, \quad m \in M \tag{8}$$

Constraint 3: This constraint is introduced to reduce travelling within a possession window. It ensures that only repair tasks that are close to each other and on the same segments are merged in a window. This is practical for operation viewpoint, because a segment can then be occupied for maintenance without completely stopping traffic on the entire line. The possibility of rerouting and redirecting will be slim if two or more segments are occupied for maintenance in the same window. Equation 9 describes this constraint, where m and m′ are two different tasks on segments $\mathrm{s_m}$ and $\mathrm{s_{m'}}$ respectively. $\mathrm{N_s}$ is the total number of segments. This constraint is handled as a quadratic constraint.

$$x_{mw} \cdot x_{m'w} = \begin{cases} 0, & s_m \neq s_{m'} \\ x_{mw} \cdot x_{m'w} \end{cases} \tag{9}$$

$$m, m' \in M, m \neq m' \quad w \in W \quad s_m, s_{m'} \in S \quad S = \{1, 2, 3..N_s\}$$

The boundary condition of the variables is defined in Eq. 10 below.

$$x_{mw} \in \{0, 1\} \quad m \in M \,\text{and}\, w \in W \tag{10}$$

However, for the alternative approach, the objective function is modified such that the start-up cost is charged per task. The new aggregated cost with the modified start-up cost is presented in Eq. 11. In addition, the quadratic constraint in Eq. 9 is replaced with a new linear constraint presented in Eq. 12. The new constraint ensures that only one task can be carried out in a window. This modification makes model to be of a linear form and can be solved as simple mixed integer linear program.

$$c_{mw} = (c_m + c_w)t_m + c_{fm} + c_{fw} + c_p \max(d_w - D_m, 0) \\ m \in M \text{ and } w \in W \tag{11}$$

$$\sum_{m \in M} x_{mw} \leq 1 \quad m \in M \, \text{and} \, w \in W \tag{12}$$

2.2 Solution

The proposed model has a linear objective function and a combination of linear and quadratic constraints. Given that the variables are binary, the model is treated as a mixed-integer quadratic constraint program (MIQCP), a special case of mixed-integer program (MIP). Gurobi optimizer was used in this study because of its accessibility and performance record on the public benchmark test set, e.g. fast solve time to feasibility and optimality.

The MIQCP is named model 1 and solved using a branch and cut algorithm that combines the advantages of a pure branch and bound scheme and the cutting planes scheme. The branch-and-bound algorithm involves systematic enumeration and exploration of a set of candidate solutions or branches that are subsets of the solution or tree and application of the lower bounding method to each candidate solution. The cutting planes tighten the formulation by removing undesirable fractional solutions during the solution process without creating additional sub-problems. A detailed description of branch and cut algorithm can be found in [15, 16]. Additional guidelines for implementation of the algorithm within the Gurobi optimizer are available in the reference literature of the optimiser [17].

The optimizer uses either the linearized outer approximation approach with the simplex algorithm or the continuous QCP relaxation approach with the barrier algorithm for both the root and other nodes in the branch and cut tree. The linearized outer approximation approach has been adopted to pre-linearizing all quadratic terms in the model. This is achieved by introducing new variables to replace the quadratic terms and introducing new constraints such that the original problem remains unchanged. The sub-problems at the tree nodes are then solved using continuous LP relaxation with simplex algorithm. Furthermore, an alternative approach named model 2 is used to model the problem using the simple mixed integer linear progam (MILP). This is done by removing the quadratic constraint and introducing a new linear constraint as explained earlier to obtain a baseline solution for comparison.

3 Case Study and Data Description

A track section in the network of the Swedish Transport Administration (Trafikverket) is considered in the case study. The line section is 130 km long single track from Kiruna to Riksgränsen. The traffic on the line is mixed, with speed of 60 km/h for loaded iron ore freight and up to 120 km/h for passenger trains. Heavy, long and slow running trains make track possession and capacity enhancement a challenging issue for the IM. In addition, the anticipated increase in traffic volume on this track section requires an efficient maintenance practice such as availability on demand. This requires that maintenance works be fitted within short track possession periods and around the demands of freight and passenger traffic.

The data used in this study include historical potential failure data, expert assessments of failure records, train movement data and cost data. The train position data recorded at some operational zones were processed to determine train-free windows that can be used for maintenance. For generating a short-term condition based maintenance schedule, 51 windows over a period of 1 month were considered usable from traffic, safety and resource-availability perspectives. The selected windows vary in duration from 1 to 3 h with an average size of approximately 1½ h and about 75 % of the windows were smaller than this average value. Furthermore, 50 maintenance tasks were selected from the historical records of potential failure; these represent the expected monthly workload. The tasks included in the monthly workload are S&C, overhead wire, rail, fastener and signal repairs, as well as ballast and sub-ballast spot tamping. Using expert experience and available data, the possession requirement of each task is estimated and the requirement varied between ¼ and 3 h, depending on the type of work and estimated extent of damage. The maintenance labour cost per hour c_m is estimated to be 217 € based on expert information and existing contracts of the infrastructure manager (IM). The possession cost per hour c_w is 178 € while the penalty cost per day c_p is estimated to be 1084 €. The fixed cost per window c_{fw} and fixed cost per task c_{fm} are 108 € each.

4 Results and Discussions

The results of using the proposed model and alternative model for efficient possession management of potential failure and deferrable failure maintenances are presented below. The short term maintenance schedules of these models are evaluated in terms of their computational times, solutions obtained, constraint violations, optimum values, number of delayed tasks, number of days with capacity reduction, average window utilisation and number of windows used.

The overall performance of the models are summarised in Table 1 and further elaborated thereafter. The two models generated optimal solutions in less than 1 min while. The optimality of the solutions of two models were proven because the gap between the best feasible solution and the incumbent optimal solution in the optimisation algorithm is lower than the set limit. None of the models violated their respective constraints, i.e. all tasks were completed, window durations were not exceeded and tasks in different segments were not scheduled together.

The performances of the models were analysed further by studying the solutions they yielded. Model 2 is an expensive approach, in the sense that it does not permit combinations of tasks on the same segment into one window. Therefore, the associated total maintenance cost associated with model 2 is higher than model 1. Looking further into the schedule generated by each model and comparing it with their respective deadlines, model 1 has the best performance with all works scheduled and no task delayed. In model 2, 4 tasks would be implemented after the deadline.

In terms of the number of days for which capacity would be affected owing to infrastructure conditions, model 2 has worse performance in comparison with model 1 with no reduction in capacity. The average window utilisation is the highest for model 1 owing to the possibility of merging maintenance tasks in a single window. Even though none of the models led to 100 % window utilisation, the proposed model (model 1) showed better performance and can even be improved if some tasks can be broken down. In terms of the number of windows utilised, model 1 utilises less windows to complete all the tasks, leaving behind four unused windows that can be used for other purposes.

In addition to the overall performance evaluation of the models given in Table 1, a breakdown of the total maintenance cost for the optimal task schedules generated by the two models is given in Fig. 2. The total direct maintenance cost C_{maint} and possession cost C_{poss} are similar for the two models because these cost elements are functions of estimated repair time and all tasks are expected to be completed in a window. However, the distinct differences between the optimality of the two models are the total penalty costs and window start-up costs. The schedule generated by model 1 has no penalty cost because no task is delayed and its window start-up cost is small because the schedule minimises the number of windows used.

The schedule generated by model 1 is shown in Fig. 3, the task are scheduled into the different windows such that the total maintenance cost is minimised. The three annotations in the figure present a short description of the task and the window for information purpose. For instance, task 1 is a maintenance work on a switch and crossing located on the first maintenance segment. It is scheduled for

Table 1 Performance evaluation of models

	Model 1	Model 2
Method	MIQCP	MILP
Computational time	<1 min	<1 min
Optimum value (€)	33267	40168
Number of delayed works	0	4
Number of affected days	0	6
Average window utilisation	87 %	82 %
Number of windows used	47	50

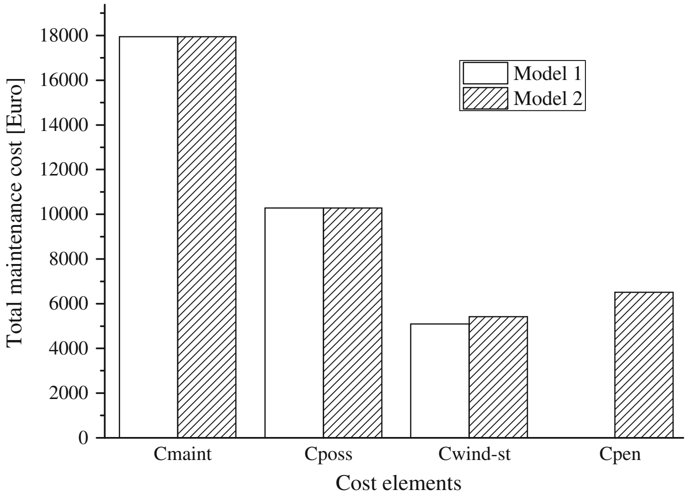


Fig. 2 Breakdown of the total maintenance cost for the proposed model and alternative model

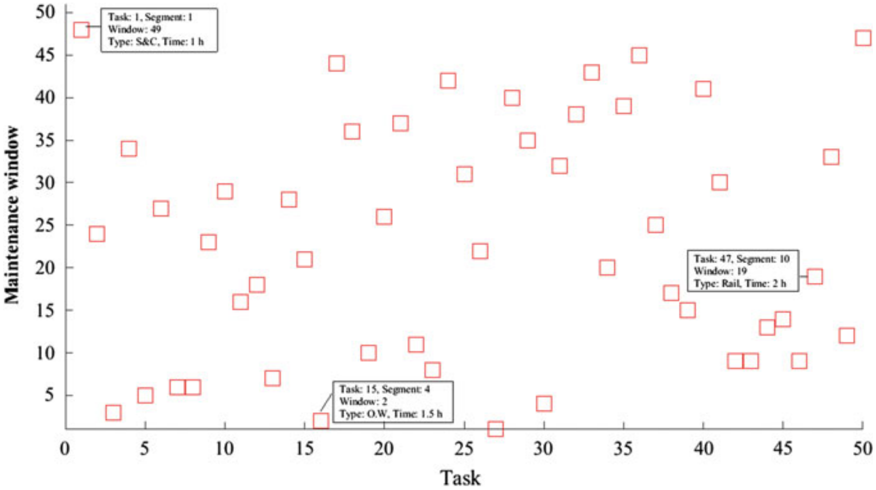
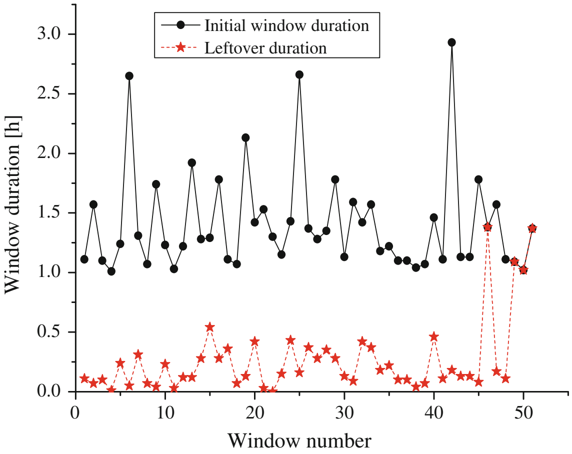


Fig. 3 Maintenance schedule for the given tasks and train-free windows

window 49 and is expected to take about 1 h besides additional 30 min that is meant for preparation, safety measures and other supporting sub-tasks.

The initial window duration and left-over time in each window for model 1 is shown in Fig. 4 for visual assessment of the possession allocation efficiency. The obviously high left-over times represent the unused windows. Approximately 80 % of the remaining windows can be considered practically unusable for repair works because they are too small to accommodate travel times and start a new task.

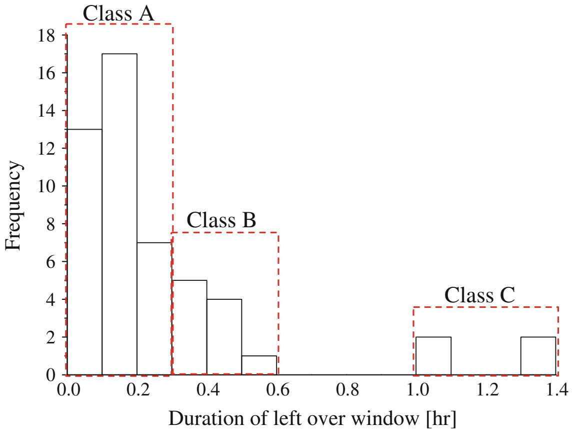
Fig. 4 Initial window duration and left time for the proposed model



Furthermore, the left over window duration can be analysed and classified, as shown in Fig. 5 for additional utilisation by other type of track works such as small inspection works and routine checks. The class A windows are efficiently used and perhaps not usable for other works if encroachment into the maintenance withdrawal time before the next train is to be avoided. The class B windows can still be used for opportunity based maintenance involving small-scale track works, routine checks or inspection on the same segment where the window time it was originally used. The class C windows are unused and can thus be used for any type of work on any segment provided other utilisation constraints are not violated.

An important aspect of the proposed approach for possession management is the analysis of the optimal schedule or reason for infeasibility. In instances where not all tasks can be scheduled in available windows owing to the number or the size of the windows, a review of the task can be conducted. For instance, the review could entail the possible break-up of some tasks into smaller chunks or removal of the less

Fig. 5 Histogram plot of the left over window durations



significant works that will be later spread over the left-over usable windows. The model can be adapted with little improvement to support other scheduling cases, including night possession with long duration, where merging of tasks in different segments is allowed within the same window. In future, the model will be extended to consider task implementation order and other technical or logistic conditions related to different tasks. It will also be extended to multiple track scenarios with additional information about the track layout from the asset information system.

5 Conclusion

This article describes the formulation of a short-term maintenance-scheduling problem to support the effective and efficient scheduling of maintenance works that are not accommodated in the long-term plan. The formulated problem focuses on reducing the sum of direct maintenance cost, possession cost, window start-up cost and penalty cost. The conclusions from the case study are as follows:

i. Possession scheduling for maintenance works can be supported with the proposed MIQCP model presented in this work.
ii. The MIQCP model with continuous LP relaxation approach gives the best performance with the lowest cost, zero task delay and zero capacity loss due to infrastructure condition for the case study.
iii. The use of maintenance windows for routine works or condition-based maintenance is a promising approach for possession management especially in corridors where complete night dedication for maintenance is impractical.

Acknowledgments The authors wish to thank Trafikverket and Luleå Railway Research Centre for financial support. We also appreciate Prof. Inge Söderqvist for his fruitful discussions.

References

1. CENELEC EN 60300-3-14 (2004) Dependability management part 3-14: application guide-maintenance and maintenance support. In: European committee for electrotechnical standardization
2. Dekker R, Budai G (2002) An overview of techniques used in planning railway infrastructure maintenance. In: IFRIM maintenance management modelling conference, Växjo, Sweden
3. CEN EN 13306 (2010) Maintenance terminology. In: European committee for standardization
4. Paragreen J (2011) High level breakdown of maintenance activities—AUTOMAIN project deliverable 2.1. In: 7th framework programme—EU research
5. Trafikverket (2007) Roadmap to Banverket's maintenance strategy (Vägledning till Banverkets underhållsstrategi BVH 800). Trafikverket
6. Trafikverket (2005) Safety inspection of railway infrastructure (BVH 807.30—Säkerhetsbesiktning av fasta anläggningar). Trafikverket

7. Stenström C, Parida A, Galar D (2014) Performance indicators of railway infrastructure. Int J Railway Technol 1(3):1–18
8. Higgins A (1998) Scheduling of railway track maintenance activities and crews. J Oper Res Soc 49:1026–1033
9. den Hertog D, van Zante-de Fokkert JI, Sjamaar SA, Beusmans R (2005) Optimal working zone division for safe track maintenance in The Netherlands. Accident Anal Prevent 37 (5):890–893
10. Cheung BSN, Chow KP, Hui LCK, Yong AMK (1999) Railway track possession assignment using constraint satisfaction. Eng Appl Artif Intell 12(5):599–611
11. Peng F, Kang S, Li X, Ouyang Y, Somani K, Acharya D (2011) A heuristic approach to the railroad track maintenance scheduling problem. Comput Aided Civil Infrastruct Eng 26 (2):129–145
12. Budai G, Huisman D, Dekker R (2006) Scheduling preventive railway maintenance activities. J Oper Res Soc 57:1035–1044
13. Zhang T, Andrews J, Wang R (2013) Optimal scheduling of track maintenance on a railway network. Qual Reliab Eng Int 29(2):285–289
14. Forsgren M, Aronsson M, Gestrelius S (2013) Maintaining tracks and traffic flow at the same time. J Rail Transp Plan Manag 3(3):111–123
15. Wolsey LA, Nemhauser GL (2014) Integer and combinatorial optimization. Wiley
16. Junger M, Liebling TM, Naddef D, Nemhauser GL, Pulleybank WR, Reinelt G, Rinaldi G, Wolsey LA (2010) 50 years of integer programming 1958–2008. Springer, Berlin
17. Gurobi optimizer reference manual (2015) http://www.gurobi.com

A Survey on Predictive Maintenance Through Big Data

Amit Patwardhan, Ajit Kumar Verma and Uday Kumar

Abstract Modern manufacturing systems use thousands of sensors retrieving information at hundreds to thousands of samples per second. The real time data being generated is mostly used for monitoring the processes and the equipment condition. Data processing techniques applied to this data to detect anomalies and thus applying preventive maintenance have been used in the industry. Currently available technologies which were developed during the last two decade for scanning the Internet and providing computational services, working at very large scale can be re-targeted to fulfil the requirements of maintenance of complex systems. These systems can support storage and processing of current as well as historical data. Ability to access and process these large data sets will lead from preventive to predictive maintenance and eventually to smart manufacturing.

Keywords Big data · Hadoop · Spark · Maintenance

1 Introduction

Industry has come a long way to reduce the waste in the production process and variability in quality and yield through various management systems. Still, production process that depend on large number of complex systems suffer from high variability and effects on the yield even after best management practices have been

A. Patwardhan (✉) · U. Kumar
Division of Operation and Maintenance, Luleå University of Technology,
Luleå, Sweden
e-mail: amit.patwardhan@ltu.se

U. Kumar
e-mail: uday.kumar@ltu.se

A.K. Verma
University College, Haugesund, Norway
e-mail: akvmanas@gmail.com

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_31

used. In order to detect and correct flaws and to have proper control over the entire production process detailed analysis of every step is required. Trend had been to collect large number of streams providing real-time data from equipment and processes, this data was used to track the system and detect the issues to apply preventive maintenance. Data was not stored for long term due to either lack of infrastructure or large costs involved in data storage. Advances in technology have made it possible not only to store huge amounts of data at a fraction of cost but also allows for processing on this data at the location of storage in a distributed manner. This changes the amount of data that can be economically stored, and the way it is processed. Facility to store and process large amount of data allows for analysis of current conditions and access to historic data helps to plot the probable progression of the system in the future. Analysis of this "Big Data" is becoming an important step towards achieving higher throughput from the production system. Data analytics as the science of analysing patterns and trends in data, used to gain insight about a system as an application of statistical and mathematical models has emerged as a powerful and effective tool to assess and improve the production process.

2 Predictive Maintenance

Predictive Maintenance primarily helps in detection of when, where and why asset failure are likely to occur. At a secondary level predictive maintenance helps to optimise inventory and helps minimise issues related to quality and reliability. It supports operations planning and in turn reduces operations cost. Successful implementation of predictive maintenance depends on accessing data from the equipment, ability to detect patterns and to be able to relate the patterns to functioning of the system.

Predictive Maintenance will not completely replace reactive maintenance as there will always be unforeseen conditions and equipment failure will occur, predictive maintenance tries to reduce the possibilities of such conditions and improve the overall reliability of the infrastructure.

3 Data Collection

Data collection is the first step towards implementing a big data system. Data collection methods depend on the access provided to the system. Methods for collecting data [1] can be through log files which are automatically generated at the data source system. Sensing of physical quantities like sound wave, voice, vibration, automobile, chemical, current, weather, pressure, temperature etc. In order to acquire network data web crawlers, word segmentation systems, packet capture systems can be used, in addition to aforementioned data acquisition methods data from scientific experiments or mobile phones may also be collected [1].

3.1 Machine to Machine (M2M) Interface

The framework of e-maintenance machine network consists of sensors, data acquisition system, communication network, analytic agents, decision-making support knowledge base, information synchronisation interface and e-business system for decision making [2]. Data being generated by an equipment and consumed by another equipment in order to use the information generated with out any human intervention forms M2M interface.

M2M covers the technologies used to implement wired or wireless communication between systems to transfer raw data from the source through a hub to a data processor to implement a monitoring or control station. Current M2M systems do not work on one to one basis instead work through specially designed protocols and require very small amount of energy to work continuously. M2M systems have been working at different levels like caller identification, automatic reading of utility meters, point of sales terminals and automobiles for many decades. M2M systems lay the foundation for using latest technologies in industrial processes.

Implementation of data collection systems, data analytics and real-time decision making has paved the way for e-maintenance and helped reduce downtime and uncertainty about the current status of the equipment and possible breakdown in the future. Proper use of available technologies will lead towards smart systems which will reduce uncertainty in the decision making process.

3.2 Internet of Things

The term Internet of Things (IoT) was first introduced by Kevin Ashton in the year 1998. It is implemented by embedding short-range mobile transceivers into a wide array of additional gadgets and everyday items, enabling new forms of communication between people and things, and between things themselves [3].

IoT forms the next layer beyond M2M interface and since it connects and caters to a larger audience the volume of data generated as compared to preceding technologies is much higher. This data being generated has more expectation to be processed where previously most of the focus was on storage of data and much less on processing it.

3.3 Data Preprocessing

As data collected from various sources will contain noise, redundancy, and inconsistency and hence it is a waste of time and resources to to store such data [1]. Analytical methods have requirements of data quality therefore, for proper data analysis per-processing data is important. The data has to be cleaned depending upon the kind of noise or artifacts present in it also depending on the kind of analysis that will be performed on the data [4].

3.3.1 Integration

Data integration refers to combination of data from different sources to create a uniform view of data. ETL (Extract, Transform and Load) is a standard process used in data processing field. Extract refers to selecting and collecting data from the source. Transformation refers to execution of a series of rules to convert the data into standard format. Loading means importing the transformed data into the target storage infrastructure.

3.3.2 Cleaning

Data cleaning improves data quality by identifying inaccurate, incomplete, or unreasonable data and then to modify or delete such data. Data cleaning applies 5 different procedures [4]: defining and determining error types, searching and identifying errors, correcting errors, documenting error examples and error types, and modifying data entry procedures to reduce future errors.

3.3.3 Redundancy Elimination

Data repetitions or surplus in a dataset is referred to as data redundancy. It can increase data transmission and storage costs, lead to data inconsistency and reduction in data reliability. Image and video data contains large amount of redundancy and it is better to use compression algorithms on such data.

3.3.4 Hindrances

Data is generally not available due to non presence of sensors in critical sections of the equipment, even when sensors are present data is only available to internal control system. In certain equipment sensor data may be transmitted by the equipment but is encoded in a proprietary format or transmitted using a proprietary protocol hence making it unreliable for use. Data is not shared due to security concerns in some cases where the data may reveal information about the job being processed or the internal workings of the equipment. Control codes may not be revealed due to chances of malicious use of equipment. If preventive maintenance methodology has been followed components may have been replace too early and hence failure data is not available in many cases. Availability of data is essential for analysis and its integrity effects the reliability of the developed models.

4 Big Data

Big data is the dataset that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time [1].

Data has to be stored and accessed in order to be processed. The storage of data may be unstructured through file systems which allows for storage of mixed type of data (text, images, audio) or structured in a database management system (DBMS). Data base management systems have been the standard for storage and retrieval of data through the use of structured query language (SQL). As the size of data has increasing to a few terabytes or even petabytes, DBMS have not been able to keep up with the requirements of data storage, retrieval and processing.

Big data is described with four critical features and have been called as the four V's of big data [5]. These are namely volume, velocity, variety and veracity.

4.1 Volume

Volume is certainly about the size of the data but does not define a specific threshold beyond which the data may be categorised as big data rather it is the data would be a candidate to be termed as big data.

4.2 Velocity

Velocity refers not only to how fast data is being received but more importantly how fast it needs to be processed to be useful. Generally a system having real-time data or or requirement of real time data analysis will be referred as to be having a high velocity. The limiting factor would be not only the volume but also the processing power and memory requirement for temporary or intermediate data during the required computation. In order to handle these factors change in hardware, software or process may be required.

4.3 Variety

Variety refers to the multiple formats of data available or being used in the process. If the entire data was purely numerical and structured, relational databased would be sufficient for the storage and processing of the data provided the data has manageable size. When the data is unstructured and may have text, documents, audio recordings, video or even social media the methodology developed for storage and processing of data through relational databased does not scale to support the requirements of the modern processes. Big data provides support for this variety in data.

4.4 Veracity

Veracity indicates the lack of integrity in data. The incoming data may be inconsistent or ambiguous due to latency, incompleteness, sensor failure, communication failure, signal sampling error and so on. If the data being used to identify trends in the system or regions of importance can not be dependent upon for correctness. This corruption of data will add up and eventually diverge the result. The data needs to be cleaned up and processes should have trigger points to reduce the accumulation of data which may divert the processing performed on the data.

5 Hadoop

Apache Hadoop [6] is an open source implementation of MapReduce [7] programming model. MapReduce model performs operations on the data in three steps [8]. The Map part is about applying a common procedure to entire data like a filter based on a certain criteria. In second step called as shuffle the data is redistributed according to the output of the map procedure and moved to different nodes which are basically processing units. Reduce is the calculation performed on different nodes. The popularity of the MapReduce procedure is due to the possibility of implementing the procedure as a multi-threaded or on large number of computers called as clusters. The parallel nature of Hadoop as a cluster also introduces fault tolerance by supporting recovery from partial failures of nodes by rescheduling the mapper or reducer job to a different node.

MapReduce operations depending on the data size may take a very large amount of time, still it is economical and faster as compared to DBMS based solutions due to non-dependency on specialised servers but the ability of the system to be setup on commodity hardware.

5.1 HDFS

Hadoop distributed file system is a distributed file system and has been designed to be highly fault-tolerant and can be set-up to work on low-cost commodity hardware [9]. HDFS has been designed to handle large datasets and can span over thousands of machines, each of which can store a part of the data. Fault-tolerance becomes an important aspect to design a system which would involve such a large number of hardware parts and hardware failure will be common. Large size of the dataset means that moving the data in order to process it will be uneconomical instead the computation to be performed should be brought over on to the same machine where the data is stored.

The HDFS cluster is divided between NameNode and the DataNode. The NameNode refers to the master server and it manages the entire file-system. The DataNode is present per node basis and handles multiple blocks of data stored on the particular system. If certain DataNode is not able to update its presence to the NameNode it will be marked as dead and no more requests are forwarded to it also the NameNode will try to maintain replicate data on multiple DataNode's as per a replication policy.

5.2 YARN

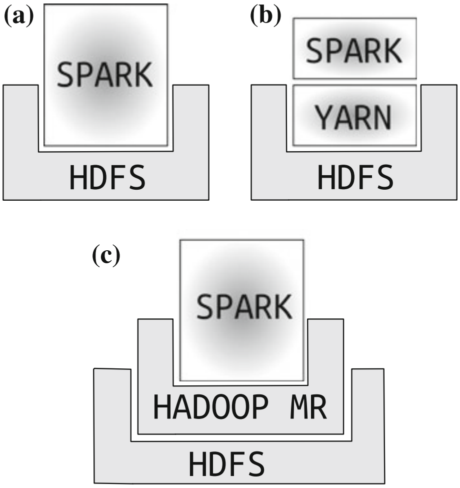
Yet Another Resource Negotiator (YARN) is basically MapReduce version two [10]. It changes the way resource management, job scheduling and monitoring is performed. A global ResourceManager and a per application ApplicantionMaster are the daemons which handle the resource management job. The ResourceManager has two parts, the scheduler which allocates resources based on resource requirements of the application and ApplicationManger executes the application specific ApplicationMaster and accepts job-submission.

5.3 Spark

Spark [11] is a fast and general processing engine compatible with Hadoop data. It can run in Hadoop clusters and it can process data in any Hadoop input format. It was designed to perform both batch processing (similar to MapReduce) and new workloads like streaming, interactive queries, and machine learning. Spark supports in-memory processing of datasets and thus improves upon the processing time.

Spark is developed in Scala language [12] and is well integration with it such that Scala can directly access and manipulate the datasets as locally available objects. Scala executes on top of Java Virtual Machine (JVM) and becomes portable to any platform which can execute the JVM with only the addition of Scala run time library. Scala can use the vast number of Java libraries available and Java code and makes it easy to utilise the investment made by the organisation in the Java stack. Spark is different than Hadoop such that is provides support for re-usability of a dataset for parallel operations, the datasets are cached in memory to reduce access latency. This feature optimise's Spark for machine learning algorithms.

Fig. 1 Spark deployed on Hadoop cluster



Spark applications are termed as drivers, these drivers may work on a single or in parallel on multiple nodes. A driver may perform an action on the data set which is basically performing a computation on the dataset and iterating over the dataset or it may perform a transformation which will create a new dataset from the old.

Spark was designed to work with Hadoop stack and has the capability to read or write to HDFS. It is used in three modes as in Fig. 1. Spark can run side by side to Hadoop MapReduce with statically allocated resources as shown in Fig. 1a. In case the transition to YARN has been made Spark can be deployed and can directly work Fig. 1b. Spark can be launched inside MapReduce with the help of “Spark in MapReduce” (SIMR) Fig. 1c. SIMR does not require administrative rights for installation and provides shell access to the user through which the user can directly access the drivers.

6 Data Analytics

Current requirement is to run systems on data-driven decisions. Scenarios and simulations should provide immediate guidance on the best action to take when disruptions occur. Ability to take optimal solutions based on complex parameters or new information is useful to take quick decisions [13].

Data analysis is the last but important step in the process of data-driven decision process. Big data differs from standard data analysis such that the size and platform in case of big data is very different as compared to a standard data analysis problems. Traditionally data analysis would be use of statistical tools, currently a large combination of analytical tools mainly clustering algorithms and correlation algorithms in addition to statistical tools are used.

Libraries for applying data analytics algorithms have been available for small data sets, currently projects [14] providing optimised implementations of required algorithms are available for use.

7 Conclusion

Big data and data analytics are powerful tools available and they are being applied to provide customised on-line advertising or even suggesting viewing options on television. Use of these tools have shown good improvements in review and user satisfaction. Application of these tools to such cases is relatively straightforward due to availability of data directly from the user.

To use the same tool set in manufacturing will require extracting the data from the equipment and processes to bring it to the data processing systems. The ability to handle this large amount of data will help to gain insight about the systems and change not only the use of predictive maintenance process but optimise overall industrial processes as well.

References

1. Chen M, Mao S, Liu Y (2014) Big data: a survey. Mob Netw Appl 19(2):171–209
2. Lee J, Ni J, Djurdjanovic D, Qiu H, Liao H (2006) Intelligent prog-nostics tools and e-maintenance. Comput Ind 57:476–489
3. Bandyopadhyay D, Sen J (2011) Internet of things: applications and challenges in technology and standardization. Wirel Pers Commun 58(1):49–69
4. Maletic JI, Marcus A (2000) Data cleansing: beyond integrity analysis. In: Proceedings of the conference on information quality, pp 200–209
5. Gantz J, Reinsel D (2011) Extracting value from chaos. IDC iView, p 112
6. The Apache Hadoop Project (2009) http://hadoop.apache.org/core/
7. Dean J, Ghemawat S (2004) MapReduce: simplified data processing on large clusters. In: OSDI, pp 137–150
8. Jiang D, Ooi BC, Shi L, Wu S (2010) The performance of MapReduce: an in-depth study. PVLDB 3(1)
9. HDFS https://hadoop.apache.org/docs/r1.2.1/hdfs-design.html-
10. YARN http://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html
11. Spark http://spark.apache.org/
12. Scala http://www.scala-lang.org/
13. LaValle S, et al. (2013) Big data, analytics and the path from insights to value. MIT Sloan Manage Rev 21
14. Mahout http://mahout.apache.org/

Part V
Probabilistic Risk and Safety Analysis

Approach for Probabilistic Safety Assessment of Accelerator

Gopika Vinod, M. Hari Prasad, G. Haridas and R.K. Singh

Abstract Probabilistic Safety Assessment revolves around identifying all the potential initiating events, developing the accident scenarios and analyzing the consequence of accident sequence. In case of accelerator, reference initiating event list is not available, which needs to be prepared based on precursor review, engineering evaluation and operating experience. Defining the consequence or risk from accelerator posed yet another major challenge. Risk in terms of absorbed dose has been proposed as one of the measure, which puts forth the hurdle of deciding the Frequency Vs Dose curve for a typical accelerator facility. There are some documents such as NUREG 1860, which proposes an F-C curve in terms of radiation dose under a techno neutral framework for consequence assessment for nuclear facilities. The paper discusses these challenges and framework developed for conducting probabilistic safety assessment of accelerators.

Keywords Probabilistic safety assessment · Accelerator · Risk ranking · F-C curve

1 Back Ground

Main goal of nuclear safety is to keep the radiation exposure from nuclear facilities to members of the public and workers as low as reasonably achievable (ALARA) during normal operational states (certainly below the limits set by the regulatory

G. Vinod (✉) · M.H. Prasad · R.K. Singh
Reactor Safety Division, BARC, Trombay, Mumbai 400 085, India
e-mail: vgopika@barc.gov.in

M.H. Prasad
e-mail: hari_m@barc.gov.in

G. Haridas
Health Physics Division, BARC, Trombay, Mumbai 400 085, India

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_32

bodies) and in the event of accident. In order to adhere to the safety goal, carrying out safety analysis has become almost mandatory for all nuclear facilities. Safety analysis of facility is aimed at the calculation of risk to the operation of facility and its comparison with other natural and industrial risks. It is based on both on deterministic and probabilistic safety analyses (PSA). In the deterministic safety analysis, design basis accidents are considered and it is shown that the engineered safety features provided to counter act such accidents result in radiation doses/risks that are acceptable. In contrast, probabilistic safety analysis (PSA) includes all possible accident scenarios and their quantification in terms of plant damage frequency and consequences. The risk (a measure of safety) in general for Nuclear facility is defined as:

Risk = Likelihood of occurrence of an accident × Consequences in terms of exposure to radioactive material release

The risk can be minimized by minimizing the accident frequency or its consequences or both. Nuclear facilities, such as accelerator also houses potential hazards. Some of the potential hazards are

- Radiation
- Heat load
- Electrical
- Ozone and
- Fire.

This paper explains how PSA is applied to hazards, in particular radiation, emanating from accelerators.

2 What Is Probabilistic Safety Assessment or PSA?

PSA provides a structured framework for evaluation of safety of any facility in a quantitative manner. PSA tries to answer the questions [1]:

(i) What can happen?
(ii) How likely is it to happen?
(iii) Given that it occurs, what are the consequences?

These questions are a starting point for defining the process of performing a quantitative risk analysis (QRA). There are essentially seven stages in QRA's implementation. These are:

1. *System description*, which is the compilation of all technical and human information needed for the analysis (including reliability data).

2. *Hazard identification*, which is a critical step in quantified risk analysis, a hazard omitted at this stage is a hazard which is not analysed.
3. *Incident enumeration*, which is the identification and tabulation of all incidents without any relevance to their importance or to the initiating event. Stages 2 and 3 may be linked together. For example, chlorine gas is a 'hazard' while its unplanned emission through a faulty valve is an 'incident'.
4. *Incident frequency estimation*, which uses likelihood of estimation models for selected incidents and evaluates frequencies. Fault tree analysis and event tree analysis are typical techniques used at this stage.
5. *Consequence estimation*, which is the methodology used to determine the potential damage or harm from specific incidents.
6. *Evaluation of consequences*, this stage is concerned with the estimation of frequency data for specified consequences. Estimates are based primarily on generic data, i.e. data abstracted from banks and various sources of historical data.
7. *Risk estimation* combines the consequences and likelihood of all incident outcomes from all selected incidents to provide a measure of risk.

Many measures of risk have been proposed and are in use, each providing a different view of a particular situation or aspect. Among these measures, perhaps most commonly used ones are those of individual risk and societal risk. Figure 1 presents the steps in risk analysis.

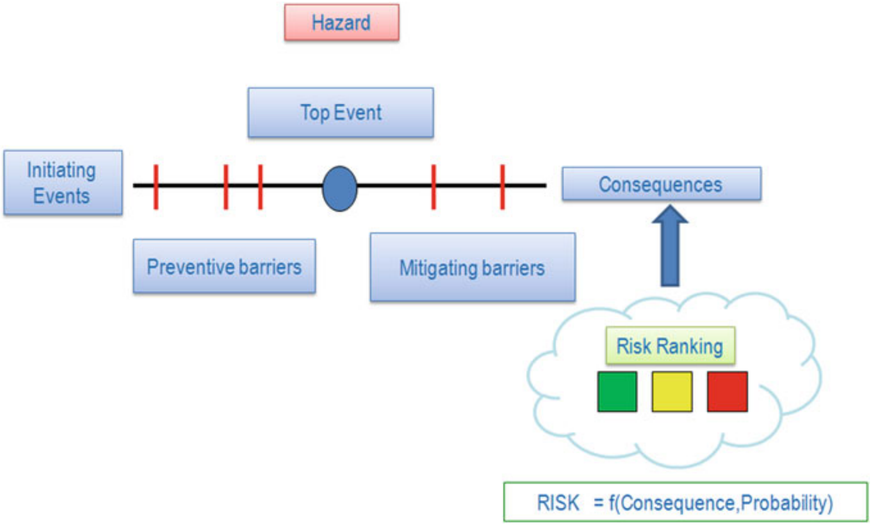


Fig. 1 Steps in risk analysis

3 Applying Probabilistic Safety Assessment for Accelerators

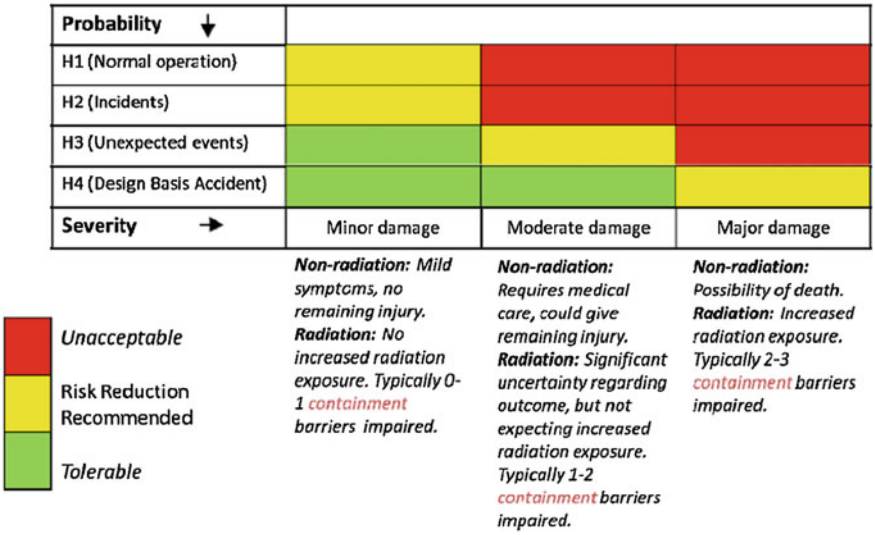
The initial task of this analysis is to gather information from the facility on design aspects and operation practices which formed the basis for development models. Information was assembled using sources such as safety analysis reports, design manuals, operating practices etc. There are different sources of hazards such as Radiation, Heat load, Electrical, Ozone and Fire exists in accelerator facility. Considering radiological hazards from electron beam accelerators, there are two types of radiation expected: synchrotron radiation and bremsstrahlung radiation. Dose from these two types of radiations are considered. In order to quantify the risk it is required to postulate all the probable initiating events which may lead to accident kind of situation if they are not properly mitigated. In preparation of list of postulated events, Precursor review, Engineering evaluation and Use of operational experience were used. Since there are no standards/documents listing postulated events from an electron accelerator is available these approaches may not be conclusive and they can even vary with design [2].

For each Postulated Initiating events, event progression and mitigation are modelled using event trees. Event tree modeling considers the procedure available to prevent the undesirable consequence in case if an event happens. The safety function can result either in tripping beam or closing the safety shutters. This activity comprises of identifying the systems involved in safety functions and probable human actions involved in event mitigation. Typically, dose received during beam dump and personnel exposures are undesired consequence considered. From the event trees it is possible to identify the various sequences that will result into accident kind of situation. In the consequence analysis one needs to find out the consequence in terms of dose absorbed for each accident sequence.

4 Decision Making from PSA

The outcome of PSA studies for these facilities is to generate a risk profile. Categorization of risk involves assessment of risk taking into consideration the Likelihood of occurrence of initiating events and severity of the desired consequence. There are two types of methods available for risk categorization such as

- qualitative risk ranking schemes
- quantitative risk ranking schemes.



Probability ↓			
H1 (Normal operation)	🟨	🟥	🟥
H2 (Incidents)	🟨	🟥	🟥
H3 (Unexpected events)	🟩	🟨	🟥
H4 (Design Basis Accident)	🟩	🟩	🟨
Severity →	Minor damage	Moderate damage	Major damage
	Non-radiation: *Mild symptcms, no remaining injury.* ***Radiation:*** *No increased radiation exposure. Typically 0-1 containment barriers impaired.*	***Non-radiation:*** *Requires medical care, could give remaining injury.* ***Radiation:*** *Significant uncertainty regarding outcome, but not expecting increased radiation exposure. Typically 1-2 containment barriers impaired.*	***Non-radiation:*** *Possibility of death.* ***Radiation:*** *Increased radiation exposure. Typically 2-3 containment barriers impaired.*

🟥 *Unacceptable*

🟨 Risk Reduction Recommended

🟩 *Tolerable*

Fig. 2 Qualitative risk matrix

4.1 Qualitative Risk Ranking Schemes

In scandpower work for risk assessment of accelerator [3], they have discussed qualitative approach using risk matrix. Three levels of risk are defined; "Unacceptable", "Risk Reduction Recommended" and "Tolerable" as shown in Fig. 2.

Unacceptable risks require risk reducing measures in order for the suggested design to be accepted. Risks Reduction Recommended require a demonstration that the suggested barriers are as effective as reasonably can be achieved considering alternatives and additions. Tolerable risks require no additional barriers, but need to be monitored, for example when design changes, to be kept at a low level.

4.2 Quantitative Risk Ranking Schemes

The risk matrix is depiction of the frequency and consequences. Quantitative treatment can also be extended to risk matrix shown in Fig. 2. The probability values can be high ($<10^{-1}$ per year), medium (10^{-2} to 10^{-1}), low (10^{-4} to 10^{-2}) or extremely low (10^{-6} to 10^{-4}). The consequences are categorized as high to extremely low based on whether the incident has serious impact on off-site and on-site.

Even the consequences can be expressed in several different units of measure which include released activity in terms of curies (or Becquerel's) of various radionuclides, health effects like early fatalities and latent cancers, and radiation

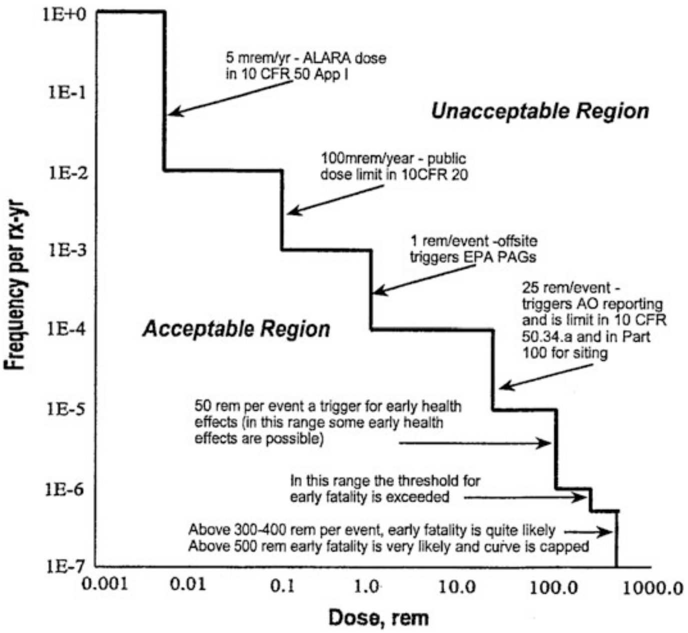


Fig. 3 F-C curve from NUREG 1860

doses (rems or sieverts). NUREG 1860 [4] has proposed an F-C curve in terms of radiation doses. One advantage of this measure is that it is based on national and international regulatory practice, e.g., NRC regulations in 10 CFR 20 and 10 CFR 50, EPA (Environmental Protection Agency) protective action guidelines, IAEA guidelines and International Commission on Radiation Protection (ICRP) recommendations. In the present study, the F-C curve given in Fig. 3 has been utilized for further analysis.

5 Case Study

Accelerator facilities have emerged as powerful tools for research, they are associated with hazards from radiation sources (bremsstrahlung radiation and neutrons), energy sources, hazardous materials etc. The likely types of failure and their consequences for the system as a whole should be taken into account. Examples include:

- Loss of access control;
- Malfunctions and failures of structures, systems and components;
- Electrical distribution faults, from localized faults to complete loss of external energy sources;

- Failure resulting from external causes;
- Failure of personnel to observe proper, safe procedures;
- Breakdown of procedures for preventing access to the facility by unauthorized persons;
- Breakdown of administrative procedures, leading to unsafe practices.

In this analysis, initiating events from beam lines are selected, which are leading to following undesirable consequences:

(i) High radiation in experimental hall due to inadvertent beam dump
(ii) Personal exposure in experimental hutch due to failure in safety barriers

Based on the discussions with designers, operators and precursor review, following initiating events were identified and are graphically represented in Table 1.

For all Postulated Initiating Events, event progression is modelled using event trees and consequence is analysed in terms of dose received from beam dump or personnel exposure. A typical event tree for "Loss of cooling" is given in Fig. 4.

Personnel Safety and Interlock System (PSIS), Safety Shutters and radiation detectors are some of the main safety systems in place to mitigate initiating events. For conducting the reliability analysis of these safety functions, detailed Failue Mode anf Effect Analysis (FMEA), preparing reliability data base, development of faul trees and conducting human reliability analysis—for preinitiator and post initiator human actions were essential.

Failure database for the identified components—(generic as well as facility specific data from events [5, 6]) is given in Table 2.

The dose received from beam dump and personnel exposure is used as the measure from consequence estimation from each postulated initiating events. F-C curve, as per NUREG 1860, is applied for communicating the risk from the events identified from the accelerator. Figure 5 shows the typical F-C curve obtained for electron beam accelerator.

Table 1 List of postulated initiating events

Undesirable consequences	
High radiation in experimental area due to inadvertent beam dump	Personal exposure in experimental hall due to failure in safety barriers
1. Loss of cooling	1. Inadvertent entry during experiment (Door not locked)
2. Vacuum degradation due to sputter ion pump failure	2. Spurious opening of safety shutter
	3. Power supply failure in PSIS
	4. Trapping of person inside
	5. experimental hutch
	6. Safety shutter fails to close during sample changing

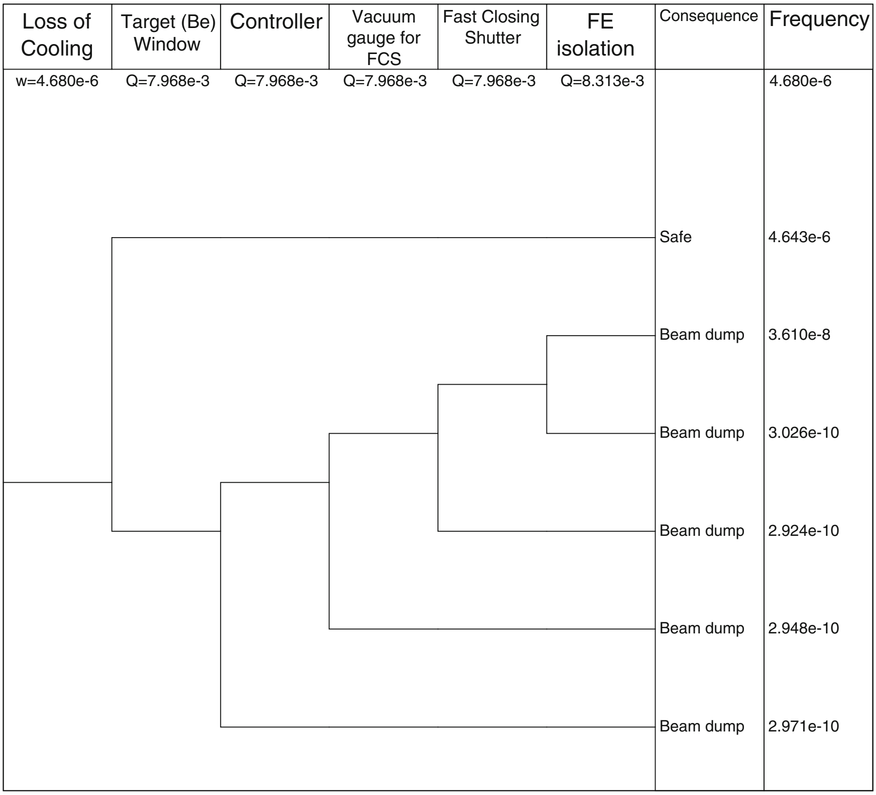


Fig. 4 Event tree for ‘loss of cooling’

Table 2 Reliability data for components in accelerator

Component name	Value	References
Flow switch	9.8E-7/h	IAEA, pg 208
Chiller failure	2.7E-6/h	IAEA, pg 135
Sputter ion pump 2	3.6E-05/h	LANSCE
Cable shorted	1.5E-07/h	ACIS (pg 56)
Switch contact fails	2E-07/h	ACIS (pg 56)
Relay fails in open position	3E-04/demand	ACIS (pg 56)
Relay contact shorted	5E-08/h	ACIS (pg 56)
Spurious signal generation	5E-08/h	Taken same as contact shorted
Radiation detector	7.4E-06/h	ACIS (pg 56)
Bellow failure	0.0046E-6/h	NPRD -91 (pg 43)
Compressor failure	3E-04/h	IAEA(pg 116)
Controller	1E-3	From manufacturer
Fast gauge	1E-3	From manufacturer
Fast cooling shutter	1E-3	From manufacturer

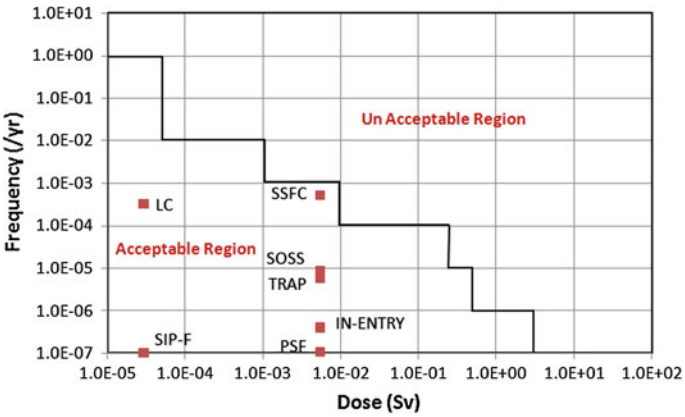


Fig. 5 Frequency verses dose curve for experimental hutch

6 Conclusion

Based on the precursor review, engineering evaluation and operating experience a list of postulated initiating events have been prepared. These events mainly lead to either beam dump or direct personal exposure in the absence of safety barriers. Event trees have been developed for all the postulated events which will depict the event progression with failure of different safety barriers. From these event trees dominating accident sequences have been identified. In order to evaluate these accident sequences, frequency of occurrence of initiating events and failure probabilities of safety barriers have been estimated by using fault tree approach. Both common cause failures and human interactions have been considered in the analysis and suitable models have been chosen to estimate the corresponding probability values. The data for the analysis have been collected from generic sources.

References

1. Ralph R (2000) Fullwood, probabilistic safety assessment in chemical and nuclear industries. Butterworth-Heinemann, Oxford
2. Advanced photon source safety assessment document (2006) APS-3.1.2.1.0, Rev. 3, work sponsored by U.S. Department of Energy
3. Risk analysis of the accelerator, instruments (2012) Scand power risk management report, 210650-R-003
4. NUREG (1860) Feasibility study for a risk-informed and performance-based regulatory structure for Future Plant Licensing, (Vol. 1), December 2007
5. Roglans-Ribas J, Nietert RE (1995) Advanced photon source LINAC/ PAR access control & interlock system reliability study, phase 2. Final report, Reactor Analysis Division, Argonne National Laboratory
6. International Atomic Energy Agency (1986) Reliability data for probabilistic safety assessment of NPP, IAEA-TeCDOC-478, Vienna

A Comparative Risk Assessment for Sites with Single and Double Units

Varun Hassija, C. Senthil Kumar, K. Velusamy and V. Balasubramaniyan

Abstract A majority of nuclear power generating sites in the world houses more than one nuclear power plant. Traditionally, a PSA is carried out to evaluate the risk associated with single unit NPP taking into account the defence in depth features and postulating combination of potential accident initiators for different hazards. The objective of PSA is to quantify risk metrics such as core damage frequency and LERF and identify weak links in the system to strengthen and ensure that safety targets are met. Post Fukushima accident, it is evident that for a site consisting of multiple units, a single reactor specific metric is not adequate and there is a need to estimate the risk arising from events affecting multiple units in the site. Our earlier work presents an approach to estimate the risk from a multi-unit nuclear power plant site. In the present paper, an attempt is made to compare the risk for a single and double unit site using the same approach. The integrated risk at a multi-unit site is estimated against various external hazards and internal events and the risk metric used is 'Site Core Damage Frequency' which is defined as the sum of all possible single and multiple combinations of core damage per site per year, with consideration of various inter-unit dependencies. The study when extended, through sensitivity analysis can form the basis to optimize the shared resources effectively at the multi-unit sites. The spin-off from such a study carried out during the design stage will provide input to decide the optimum number of units at a site, the optimal distance between two units, etc.

V. Hassija (✉) · K. Velusamy
Reactor Design Group, Indira Gandhi Centre for Atomic Research (IGCAR), Kalpakkam,
Tamil Nadu 603102, India
e-mail: varunhassija@gmail.com; varunhassija@igcar.gov.in

K. Velusamy
e-mail: kvelu@igcar.gov.in

C. Senthil Kumar · V. Balasubramaniyan
Safety Research Institute (SRI), Atomic Energy Regulatory Board (AERB), Kalpakkam,
Tamil Nadu 603102, India
e-mail: cskumar@igcar.gov.in

V. Balasubramaniyan
e-mail: bala@igcar.gov.in

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_33

Keywords Multi unit nuclear site · Site core damage frequency · Twin unit site · External hazards · Shared resources

1 Introduction

The Fukushima disaster in 2011 has highlighted the need of risk assessment for a multi unit nuclear power plant (NPP) site. Moreover, most of the nuclear power generating sites in the world are housing more than one NPP. Hence it is imperetive to estimate the risk from such sites against various potential hazards like earthquake, tsunami, flood, etc.

Probabilistic Safety Assessment (PSA) is a systematic and comprehensive methodology to evaluate risk. It quantifies the risk metrics such as Core Damage Frequency (CDF) and Large Early Release Frequency (LERF) and identifies the weak links in the plant. This helps in further enhancement of safety to meet the stipulated regulatory requirements.

However, the accident at Fukushima has shown the limitation of the risk metric CDF to capture the risk for a multi unit NPP site. To overcome this, our earlier work [3] discussed yet another risk metric, viz., Site Core Damage Frequency (SCDF) for sites having more than one NPP. SCDF is defined as the sum of all possible single and multiple combinations of core damage per site per year. An approach to quantify SCDF for a multi-unit NPP site taking into account both internal events and external hazards was presented. Modelling of inter-unit dependencies, shared resources, etc. was also demonstrated in the study.

In the present paper, the same approach is used to compare the risk for a single and twin unit site. The fact that most of the NPPs in the world are situated at a twin unit sites is apparent from the data provided in Table 1 [5]. It is also observed that about 50 % of the NPPs in the world are located either at single or twin unit sites.

Table 1 Proportion of nuclear power plants at various sites in the world

No of units at a site	NPP %
1	13.59
2	35.02
3	10.37
4	27.65
5	0.00
6	8.29
7	3.23
8	1.84

2 Description of the Two Sites

The two sites under study is assumed to house identical design of Pressurised Heavy Water Reactor (PHWR). The main engineered safety systems of the NPP are:

Reactor Protection System: Each unit is equipped with two diverse and independent shutdown systems:

- Primary Shutdown System: The system consists of mechanical shutoff rods which get quickly inserted in the reactor core following a reactor trip signal under the action of gravity and initially assisted by a spring thrust [1].
- Secondary Shutdown System: It consists of vertical empty tubes located in the reactor core into which liquid poison is injected whenever the system is called upon due to a trip signal [1].

Shutdown Cooling System: The shutdown cooling system of the NPP is comprised of two cooling trains. The trains take the decay heat away from the reactore core. Each train is having one shutdown cooling pump (SDCP) and one shutdown heat exchanger (SDHX) which dissipates its heat to the process water. Emergency process sea water pumps are used to circulate process sea water through the process sea water heat exchangers in once through mode to vent out the heat to the sea. A typical PHWR is equipped with two dedicated process sea water heat exchangers and three emergency process sea water pumps. Successful operation of any one heat exchanger and pump is sufficient to meet the post shutdown heat loads.

Emergency Core Cooling System: This system is deployed to remove the decay heat from the core of the reactor in order to mitigate the consequences of Loss of Coolant Accident (LOCA) in the rare event of break in primary circuit pressure boundary. The emergency core cooling system (ECCS) operates in two phases. In the first phase of operation, high pressure heavy water from accumulators is injected into the reactor core via headers whereas in the second phase (recirculation phase), water is taken up from the suppression pool and is injected into the reactor after passing it through the ECCS Heat Exchangers. The ECCS Heat Exchangers transfers its heat to the process sea water heat exchanger with the help of process water and is vent out to the sea with the help of emergency process sea water pumps.

Apart from these engineered safety systems the plant is also equipped with other safety support equipments, systems and infrastructure. The configuration of these support systems is site specific as sharing for them takes place between the units at a multi unit site. The description of such systems and their structure/configuration in a typical Indian multi-unit site is given below:

Diesel Engines: These are meant for fire water injection. Successful operation of one diesel engine will ensure sufficient supply of water for the decay heat removal of maximum two units.

Diesel Generators: These are deployed to take the emergency loads of the NPP like Decay Heat Removal (DHR), emergency lighting, egress lighting system lamps and

Table 2 Various systems, structures and components (SSC)/safety support systems for the two sites

Systems, structures and components (SSC)/safety support systems	Success criteria	
	Single unit site	Twin unit site
Diesel generators	1/3:S	2/5:S
Diesel engines	1/2:S	1/4:S
Switchyard buses	1/2:S	2/3:S
Compressors	1/2:S	2/4:S
Sea water intake tunnel	1	1

for charging AC UPS System and DC control power supply systems. Operation of one diesel generator is sufficient for meeting all the emergency loads of a single unit.

Sea Water Pump house: The sea water pump house deployed at the site houses condenser cooling water, process sea water and emergency process sea water pumps for both the units. The five condenser cooling water pumps and the three process sea water pumps which are installed for each NPP are driven by class 4 power supply. But the three dedicated emergency process sea water pumps are driven by class 3 power supply and availability of any one of them will ensure sufficient supply of water for DHR of a single unit.

Switchyard: The NPP is connected to the electrical grid system for class 4 power through a switchyard which also facilitates export of plant generated electric power to the grid.

Sea Water Intake tunnel: This tunnel is made to provide the sea water to the NPPs and serves as the ultimate heat sink.

Compressed Air System: The site has a compressed air station for supplying compressed air to the NPPs. Operation of one compressor ensures sufficient supply of all air (Instrument, Service and Mask air) for a single unit.

The configuration of the critical infrastructure for the two sites is described in Table 2 and the schematic of the twin unit site is shown in Fig. 1.

3 Multi Unit Risk Assessment

3.1 The Methodology

In this approach, external hazards and internal events are categorized as definite and conditional [2, 4, 6, 8, 10] The hazards that will always affect multiple units are known as definite hazards and those which only under certain circumstances affect multiple units are to be called as conditional hazards. After the initiating events for external hazards and internal events are identified and categorized, event tree/fault tree models are developed for each hazard category for the subsequent analysis. The key issues which need to be addressed while modelling event trees and fault trees

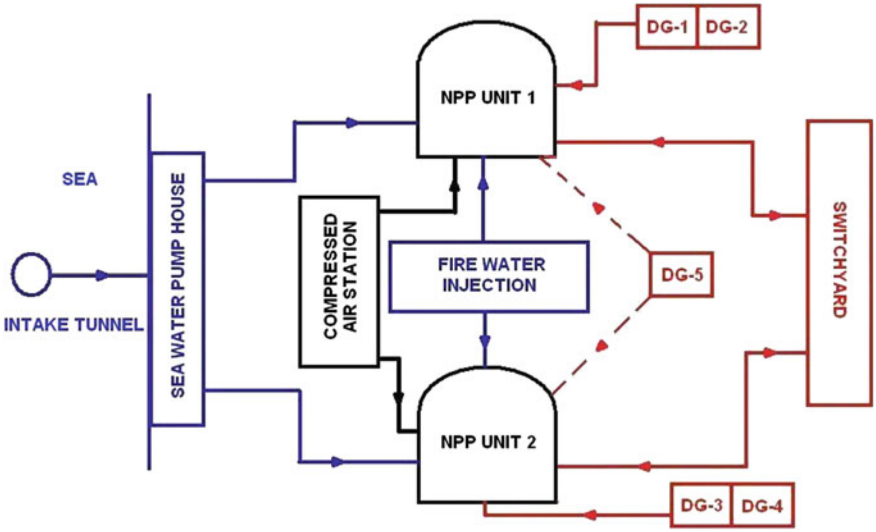


Fig. 1 Schematic of twin unit PHWR site

for the multi-unit site safety assessment are identified [3, 8]. These key issues are classified as shared systems or connections, identical components, human dependencies, proximity dependencies, mission time and cliff edge effects. These issues account for the various dependencies [8] which exists between the units owing to shared physical links, similarity in the design, installation and operational approach for a component/system, same or related environment of positioning the systems and various human interactions.

3.2 Hazards, Initiating Events and Key Issues Modeled

Hazards, initiating events and key issues modeled are listed in Table 3. The list is not comprehensive as only selected representative events considered in the study for demonstration of the methodology is shown. The details of the key issues are discussed below:

Mission Time: The mission time for accident sequences of various hazards is decided based on the nature and severity of the hazard. A mission time of 72 h is taken for external hazards such as tsunami, earthquake, clogging, etc. whereas mission time of 24 h is selected for all internal events.

Cliff Edge Effect: The cliff edge effect has been modeled for both the sites while estimation of risk from the tsunami hazard. In this, the fragility of vulnerable and unprotected component is taken as unity during occurrence of tsunami above the design limit. In this analysis, a design height of 10 m is used.

Table 3 Hazards, initiating events and key issues modelled

Category of hazard		Hazard	Initiating event	Key issues modeled	
Single unit	Twin unit			Single unit	Twin unit
External hazards	Definite external hazards (DEH)	Earthquakes	Loss of offsite power	• Mission time	• Mission time
					• Proximity dependency
					• Shared SSC
				• Proximity dependencies	• Identical components
		Tsunami	Loss of offsite power	• Cliff edge effect	• Cliff edge effect
					• Mission time
					• Proximity dependency
					• Shared SSC
				• Mission time	• Identical components
				• Proximity dependencies	
	Conditional external hazards (CEH)	Clogging in intake tunnel	Loss of ultimate heat sink	• Mission time	• Mission time
					• Proximity dependency
				• Proximity dependencies	• Shared SSC
Internal events	Definite internal initiating events (DIIE)	–	Loss of offsite power	• Mission time	• Mission time
				• Proximity dependencies	• Proximity dependency
					• Shared SSC
	Conditional internal initiating events (CIIE)	–	Loss of instrument air		• Mission time
					• Proximity dependency
					• Shared SSC
	Internal independent events (IIE)	–	Primary-LOCA		• Mission time
					• Proximity dependency
					• Shared SSC
		–	TOPA/LORA		• Mission time
					• Proximity dependency
					• Shared SSC

Shared Components: Sharing of the components at the twin unit site exists in two ways.

(a) **Same SSC shared between both the units**: This sharing exists for diesel engines and compressors. Here the components are assigned the same name and they are appearing as common components in the fault trees/event trees of both the units.

(b) **Standby System Sharing**: One diesel generator is shared by both the units at the twin unit site. Sharing of resource in a multi-unit site is modeled by assigning preference probability of the component/system for a particular unit [8]. It is assumed that preference of the common component/system (in this study DG5) will be given to unit 1 with preference probability of 0.75. The unavailability of the common component/system is then suitably estimated for the individual units with appropriate preference probability. For e.g. DG5 unavailability for unit 1 (DG_{u1}) is estimated as

$$DG_{U1} = (1 - Pf_{u1}) + (Pf_{u1} * P_{DG5})$$

and DG5 unavailability for unit 2 (DG_{u2}) is

$$DG_{U2} = Pf_{u1} + (1 - Pf_{u1}) * P_{DG5}$$

where Pf_{u1} is preference probability for unit 1 and P_{DG5} is the probability of DG5failure.

Identical Components: The identical components in both the units like shutdown cooling pumps, emergency core cooling pumps, diesel generator and emergency process sea water pumps are grouped under common cause failures (CCF) for which beta factor model is used. The grouping of the identical components and the value of the beta factor is based on the nature and severity of the hazard. In our study, simultaneous failure of identical components for both the units is considered only for DEH.

Proximity Dependencies: The components which share the same operating environment or failure of components that can induce failure of the other nearby components are grouped together under CCF and beta factor is used. This modeling has been done for emergency process sea water pumps.

3.3 Estimation of Component Unavailability

The unavailability of each of the component for both the sites is estimated as per the type and severity of the hazard.

3.3.1 Seismic Fragility

For earthquakes, mean fragility of the components is used for estimating the seismic risk. The mean fragility of the components is estimated using [7]:

$$P(A \le a) = \varphi\left(\frac{1}{\beta_c}\ln\left(\frac{a}{A_m}\right)\right)$$

where A_m is the median ground acceleration capacity, 'a' is the PGA value for which probability of failure, P is determined and

$$\beta_c = \sqrt{\left(\beta_r^2 + \beta_u^2\right)}$$

3.3.2 Tsunami Fragility

During Tsunami, components failures are classified in two categories:

(a) Failure of the component due to submergence: In this case, if the tsunami height is less than the component height then zero is assigned as the fragility of the component. If the tsunami height is equal to component height then the component fragility is taken as 0.1 [9] and for the case when the tsunami exceeds the component height, its fragility is taken to as unity.

(b) Failure of the component due to loss of support structure: In this case for a given tsunami height, the equipment failure probability is taken as the fragility of the support structure.

3.3.3 Clogging of the Intake Tunnel

Although the phenomeon of clogging of intake tunnel is external, the components of the NPP may become unavailable only due to internal random failures during that time. Hence, internal event data is used for this hazard.

3.3.4 Internal Events

In the case of internal events, the only cause of unavailability of the component is random failure. Typical models for internal random failures are used to estimate the unavailability of the components.

It is to be noted that in case of external hazards like earthquake and tsunami, the components may also become unavailable on account of internal random failures. Hence, unavailability due to internal random failures is also added to the seismic/tsunami fragility to obtain the total unavailability which is finally used for estimation of the risk due to these hazards.

3.4 Estimation of Site Core Damage Frequency

The risk for a single unit site is the total CDF obtained from internal events and external hazards whereas the risk for twin unit site is obtained as SCDF by summing the risk from all the categories of external hazards and internal events. SCDF is expressed as:

$$\text{Site CDF for Single Unit} = \sum_{i=1}^{2}\sum_{j=1}^{m} CDF(i,j)$$

$$\text{Site CDF for Multi unit } = \sum_{i=1}^{5}\sum_{j=1}^{m}\sum_{k=1}^{n} CDF(i,j,k)$$

where

i denote the category of hazard or event
j denote the type of hazard in the ith category
m denote the total number of types of hazard in the ith category
k denote the number of simultaneous core damages
n denotes the number of units at the site

Therefore, CDF (i, j, k) denotes the frequency of k number of simultaneous core damages due to j type of hazard in the ith category;

For a single unit site, i denote external and internal event whereas for multi-unit site,

i = 1 refers to definite external hazards for the site
i = 2 refers to conditional external hazards for the site
i = 3 refers to definite internal events for the site
i = 4 refers to conditional internal events for the site
i = 5 refers to internal independent events considering for all units

4 Results And Discussions

Site CDF for single unit and twin unit site is estimated using the integrated approach developed earlier [3]. The approach is also used to estimate the single and double core damage frequency for the twin unit site.

Generic component failure data available in the literature is used for the analysis. It is seen that in both the single unit and twin unit site, risk from internal events contribute more to the site CDF as compared to risk from external hazards. Risk from external hazard for the single unit site is 1.21E-05/yr and it is 3.87E-05/yr for the twin unit site. Risk from internal events for the single unit site is 1.65E-05/yr and for the twin unit site is 1.39E-04/yr. The increase in risk from internal and external hazard in a twin unit site is attributable to the sharing of a common DG between two units at the site. However, the contribution of external hazard to the overall risk is 42 and 22 % for single unit and twin unit site respectively (Figs. 2 and 3).

In twin unit site, the risk due to internal events is contributed significantly by definite internal events. The contribution from internal independent events and conditional internal events is negligible.

For the twin unit site the risk of double core damage (1.18E-05/yr) is found to be significantly lower than single core damage (1.66E-04/yr) and is depicted in Fig. 4.

The site CDF for the single unit site is 2.86E-05/yr whereas it is 1.78E-04/yr for the twin unit site (Fig. 5). For the system description considered in this study, sharing of the safety critical equipment viz., DG5 by both the units is major reason for the significant increase in SCDF in twin unit site.

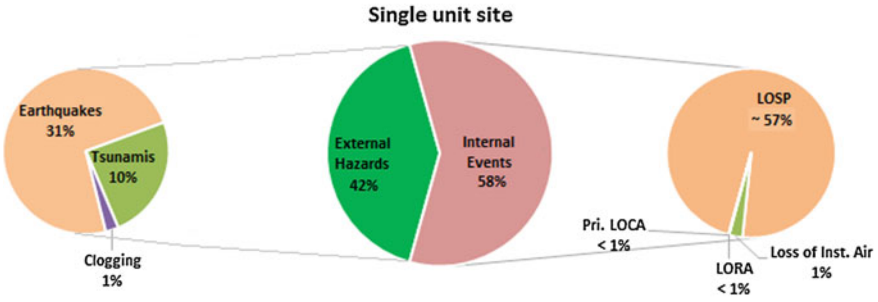


Fig. 2 Risk from single unit site

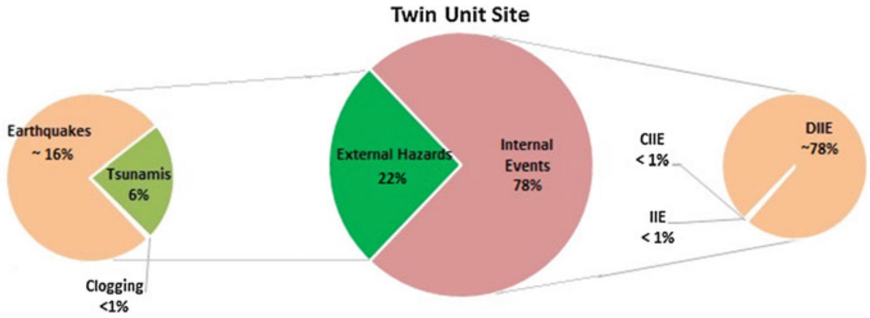


Fig. 3 Risk from twin unit site

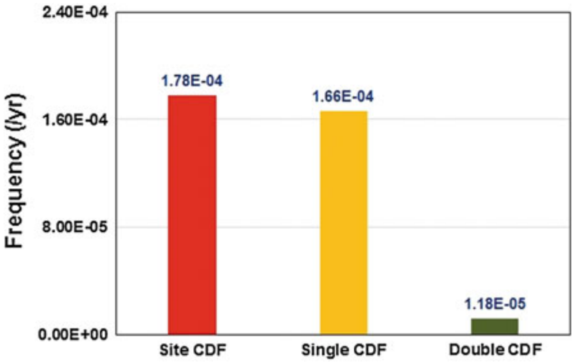


Fig. 4 Breakup of SCDF for twin unit site

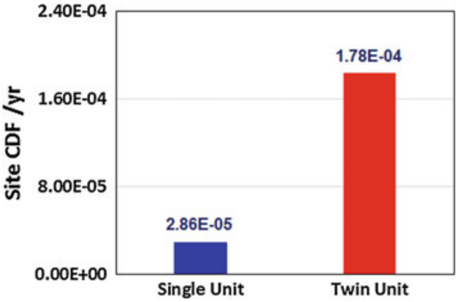


Fig. 5 Comparison of site CDF

5 Conclusions

A pragmatic approach to estimate the risk from a multi-unit nuclear power plant site is demonstrated on single and twin unit site. The method is capable of estimating the frequency of single and multiple core damage for a multi unit site against both external and internal hazards. Various key issues applicable for a multiple unit NPP site like initiating events, shared connections, identical components, proximity dependencies, cliff edge effects and mission time are accounted.

The multi-unit risk assessment methodology demonstrated with a case study reveals that the increase in SCDF in twin unit site is mainly due to a shared resource. It clearly highlights that such study will help in identification of critical structures, systems and components (SSCs) that are crucial for safety in such sites which is otherwise overlooked by carrying out only an individual unit risk assessment. Further, sensitivity analysis can form the basis to optimize the shared resources effectively. The spin-off from such a study carried out during the design stage will provide valuable inputs such as optimum number of units at a site, the optimal distance between two units and configuration of shared systems to minimize the risk in a multi-unit site.

References

1. Bajaj SS, Gore AR (2005) The Indian PHWR. Nucl Eng Des 236(7–8):701–722
2. Fleming KN (2005). On the issue of integrated risk—a PRA practitioners perspectives. In: Proceedings of the ANS international topical meeting on probabilistic safety analysis, San Francisco, CA
3. Hassija V, Senthil KC, Velusamy K (2014) Probabilistic safety assessment of multi-unit nuclear power plant sites—an integrated approach. J Loss Prev Process Ind 32:52–62
4. IAEA (2011) A methodology to assess the safety vulnerabilities of nuclear power plants against site specific extreme natural hazards
5. IAEA (2014) Nuclear power reactors in the world. Vienna
6. IAEA-TECDOC-1341 (2003) Extreme external events in the design and assessment of nuclear power plants
7. Reed JW, Kennedy RP (1994) Methodology for developing seismic fragilities. Final Report TR-103959, EPRI
8. Schroer S, Modarres M (2013) An event classification schema for evaluating site risk in a multi-unit nuclear power plant probabilistic risk assessment. Reliab Eng Syst Saf 117:40–51
9. Takeshi M (2011) Discussions of Fukushima nuclear power plant accidents by a viewpoint of PSA. Nucl Saf Simul 2(3):226–235
10. Zerger B, Ramos MM, Veira MP (2013) European clearinghouse: report on external hazard related events at NPPs, Joint Research Centre of the European Commission

Numerical Analysis of a Railway Compartment Fire

Anwar Enbaya, Taimoor Asim, Rakesh Mishra and Raj B.K.N. Rao

Abstract Trains are considered to be the safest on-land transportation means for both passengers and cargo. Train accidents have been mainly disastrous, especially in case of fire, where the consequences are extensive loss of life and goods. The fire would generate smoke and heat which would spread quickly inside the railway compartments. Both heat and smoke are the primary reasons of casualties in a train. This study has been carried out to perform numerical analysis of fire characteristics in a railway compartment using commercial Computational Fluid Dynamics code ANSYS. Non-premixed combustion model has been used to simulate a fire scenario within a railway compartment, while Shear Stress Transport k-ω turbulence model has been used to accurately predict the hot air turbulence parameters within the compartment. The walls of the compartment have been modelled as no-slip stationary adiabatic walls, as is observed in real life conditions. Carbon dioxide concentration (CO_2), temperature distribution and air flow velocity within the railway compartment has been monitored. It has been observed that the smoke above the fire source flows to both sides of the compartment. The highest temperature zone is located downstream the fire source, and gradually decreased with the increase in the distance from the fire source. Hence, CFD can be used as an effective tool in order to analyse the evolution of fire in railway compartments with reasonable accuracy. The paper also briefly discusses the topical reliability issues.

A. Enbaya · R. Mishra
University of Huddersfield, Huddersfield HD1 3DH, UK
e-mail: U0975950@hud.ac.uk

R. Mishra
e-mail: r.mishra@hud.ac.uk

T. Asim (✉)
School of Computing and Energy, University of Huddersfield,
Huddersfield HD1 3DH, UK
e-mail: t.asim@hud.ac.uk

R.B.K.N. Rao
COMADEM International, Birmingham B29 6DA, UK
e-mail: rajbknrao@btinternet.com

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_34

Keywords Computational fluid dynamics (CFD) · ANSYS · Railway compartment · Fire modelling · CO_2 concentration

1 Introduction

In the event of a train fire, the fire itself does not present the first danger to passengers. Instead, smoke from the fire is the primary danger. The inhalation of smoke causes the majority of fire-related injuries due to its emission of toxic gases. Smoke in the air can make it difficult for passengers to see exit doors clearly. Consequently, train passengers would be at serious risk of severe injury or even death, if the fire and smoke are allowed to become worse. It can be difficult to deal with train fires effectively, especially if there are many passengers on board [1]. The present study explores the flow and generation mechanism of temperature distribution and smoke in train fires.

A successful fire safety design can save lives. It is stated that large train fires can have severe consequences [2]. Cost is a primary consideration when designing a fire safety strategy. Fire size and occurrence can be lowered with higher levels of knowledge about the fire. It has been argued that designers are unable to fully estimate fires and they do not have enough knowledge of the ways in which fires behaves in the context of trains [3]. Air velocity is largely influenced by Heat Release Rate (HRR), which is generated from the train that has caught fire. Furthermore, the emergency tunnel ventilation system's performance is also controlled by this parameter. The type and amount of flammable materials within the train carriage, the characteristics of the train carriage (i.e. size, doors, windows, etc.), and the carriage construction type, determine HRR values [4]. HRR has the largest influence on how serious the fire becomes [5].

The aims of fire safety processes are to continuously improve and develop new systems that are responsive in case of fire emergencies. Currently, a new emergency response system is being explored to unite forecast and live sensor monitoring of fire development. The estimation of fire dynamics in the compartment is envisioned which will infer a paradigm change in the reaction to traumas by offering the fire service with vital evidence about the fire well ahead of time [6].

The precise forecast of the spread of smoke (poisonous gases) and distributions of temperature and velocity is vital for the scheme of detection of fire, and safety methods. It is also significant in offering testimony for the efficiency of precautionary actions, and controls the ventilation for smoke. The protocols for fire protection have to be employed while designing structures. In specific cases, for instance large public structures, museums, train stations, concert halls and tunnels, the fire protection protocols are very much significant in the case of emergency. Using these protocols, removal of the smoke generated is carried out using the instated devices throughout the time needed for the process of evacuation [7].

Modern fluid power systems are becoming complex and globally distributed. A number of issues are emerging which affect its reliability in one form or another. Some of these issues are common cause failures, reliability of computer codes and software, cyber security and risks, environmental safety issues, obsolescence issues, human factors such as behaviour, decision making, modelling and simulation issues, maintenance related issues etc. Reliability and robustness are critical and despite the advanced modelling and simulation techniques employed, many companies are finding that the operating environmental conditions are much harsher than predicted.

Uncertainties, Unreliability and Unavailability are closely related to each other and should be seriously treated as totally unacceptable by everyone right from the start. Historical evidences and plenty of literature exist to reveal the cost of unreliability and bad decision making in all walks of life. Its long-term consequences on health, wealth, quality of life and sustained prosperity of individuals and nations are recorded in histories of nations. Any number of multi-dimensional warranties, guarantees, laws and byelaws, etc. will never solve the problem of unreliability. Continuous awareness of the cost of unreliability and its dire short-term and long-term consequences should be effectively disseminated at all levels as a number one priority by all responsible people irrespective cast, creed, colour and religion. A number of case studies and bench marking studies exist that should be brought to light. It is time to initiate action research and action learning programs. Interdisciplinary research provides useful answers to many unanswered issues. Smart Integration, collaboration and proactive activities between industries, academia and professional organizations should be accelerated.

It is essential that effective systems for dealing with fire are created and employed in order to negate the risk of fire in passenger carriages, night carriages, restaurant carriages, power electronic parts and engine parts. Therefore, rail sectors require effective methods for investigating the spread of fire and the effectiveness of fire fighting strategies. However, the understanding of the evolution of fire in railway compartment needs to be understood in detail first, and hence Computational Fluid Dynamics (CFD) based analysis has been used in the present study to provide assistance to rail engineers when selecting the optimal fire fighting system setup [8, 9]. Smoke's dispersion process in space and time, temperature and velocity variations and their effects on the train's emergency evacuation systems have been investigated in the present study.

2 Numerical Modelling

The numerical analysis of a fire in a railway compartment has been carried out with the aid of a commercial CFD code ANSYS. ANSYS code comprises of the physical models involving heat transfer, turbulent flows, chemical mixing, reacting flows, multiphase flows and combustion. Finite-volume method is used by ANSYS code to numerically solve the equations that govern a fluid [10]. The carbon dioxide concentration and temperature distribution in the railway compartment are being analysed in detail. Three dimensional partial differential equations for the conservation of energy; momentum and mass are iteratively solved over a time of 360 s, with a time step size of 1 s.

ANSYS's non-premixed combustion model has been implemented to simulate non-spreading fire in a railway compartment. Non-premixed combustion model consist of the solution of transport equations for one or two conserved scalars (the mixture fractions). Equations for individual species are not solved. Instead, species concentrations are derived from the predicted mixture fraction fields. The thermochemistry calculations are pre-processed and then tabulated for look-up in ANSYS. Interaction of turbulence and chemistry is accounted for with an assumed-shape Probability Density Function (PDF) [11, 12].

Combustion includes chemical reactions with the oxygen around it dragging air into the fire and generating hot gases that travel upwards. For a methane fire the reaction for complete combustion is:

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$

Incomplete combustion refers to a lack of air. However, in well ventilated conditions, the reaction follows the stoichiometry for complete combustion, so the quantity of carbon monoxide (CO) produced is negligible, hence is the case in present study.

2.1 Geometry of the Flow Domain

The dimensions of the computational domain of a railway compartment are 20 m × 2.7 m × 2.4 m, which correspond to a conventional train compartment size in the UK, as shown in Fig. 1. The fire has been numerically initiated in the centre of the compartment using a rectangular methane burner, having a surface area of

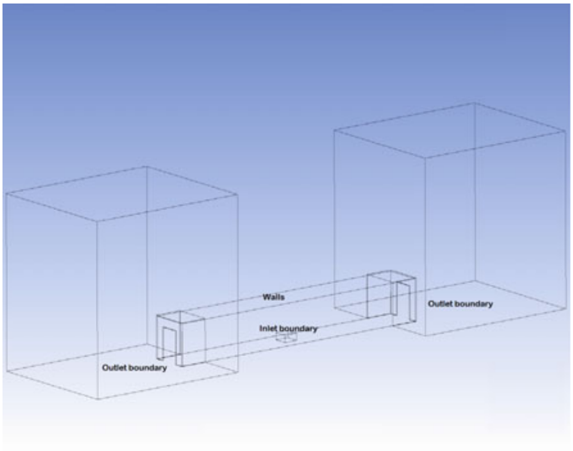


Fig. 1 The geometry of the railway compartment

1 m^2 and height of 0.4 m. The fuel flow rate specified corresponds to a heat flux of 350 kW. Ventilation is provided by two doors of 1.9 m height × 1.4 m width, one at each end, and assuming that the doors would open once the fire has been initiated.

To model the airflow through the open doors properly, and to minimize the effects of the boundaries on the fire development within the compartment, the outlet boundary of the computational domain has been extended outside both doors by 10 m × 10 m × 10 m to include a region outside the railway compartment.

2.2 *Mesh Sensitivity Analysis*

In the CFD process, the quality of meshing plays a vital role. Hence, a mesh independent analysis has been performed to confirm the precision of the results, and to identify the most effective mesh sizing in order to achieve an appropriate mesh discretisation. The mesh has been created for the compartment using the Cutcell method. Figure 2 shows the meshing of the flow domain.

The numerical simulation is run using different mesh sizes. The first mesh comprises of 346,646 elements, the second mesh of 672,308 elements, and the third mesh of 709,906 elements. The average temperature distribution within the compartment from these three simulations is compared in Fig. 3.

Fig. 2 The mesh

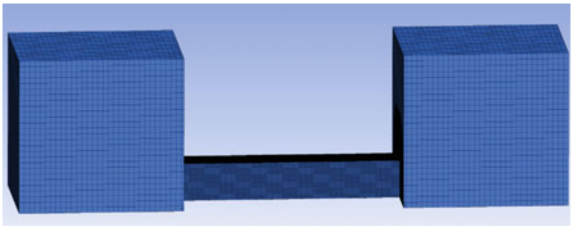
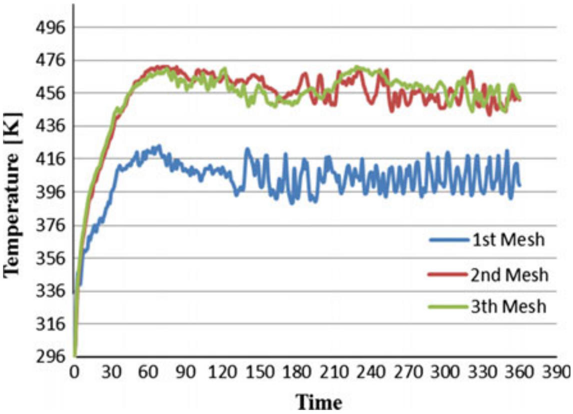


Fig. 3 Mesh independence analysis



It has been observed that the average temperature distribution within the compartment is well predicted by both the second and the third meshes, and the obtained results do not show significant changes. Furthermore, it is obvious that the simulation using the third mesh is less unstable. Therefore, this study employs the third mesh (comprising of 709,906 mesh elements) to investigate the air flow and temperature distribution within the railway compartment.

2.3 Solver Setup

Software Reliability is defined as: the probability of failure-free software operation for a specified period of time in a specified environment. Although Software Reliability is defined as a probabilistic function, and comes with the notion of time, we must note that unlike traditional Hardware Reliability, Software Reliability is not a direct function of time. Physical assets may will age with time and usage, but software will not age or wear-out during its ‘life cycle’. Software do fail due to many reasons such as: errors, ambiguities, oversights or misinterpretation of the specification that the software is supposed to satisfy, carelessness or incompetence in writing code, inadequate testing, incorrect or unexpected usage of the software or other unforeseen problems. Hardware faults are mostly physical faults, while software faults are (intentionally or unintentionally) human—induced design faults, which are much harder to visualize, classify, detect, and correct. Furthermore, design faults are closely related to fuzzy human factors, which we yet to fully understand. The quality and reliability of software will not change once it is uploaded and start running. Trying to achieve higher reliability by not knowing the root causes will hinder progress and is very costly.

Since the fire scenario is associated with chemical reactions, non-premixed combustion model has been implemented. The energy equation has been iteratively solved in order to predict the variations in the temperature. The solver settings and essential boundary conditions are summarised in Table 1.

Table 1 Solver setup and boundary conditions

Parameter	Description
Mode	Transient
Solver	Pressure based
Turbulence model	Shear stress transport k-ω
Inlet	Mass flow inlet
Outlet	Pressure outlet
Walls	Adiabatic—no slip

3 Results and Discussion

The temporal distribution of temperature within the railway compartment is depicted in Fig. 4, where Fig. 4a represents the scenario after 8 s of fire eruption, while Fig. 4b represents the scenario after 360 s of fire eruption within the compartment.

It can be seen that as the fire erupts within the railway compartment, because of its higher temperature (and lesser density), the hot air from the fire travels against the gravitational force, until it comes in contact with the ceiling of the compartment. Then it starts to spread outwards in both directions until it escapes out from the evacuation doors, and into the environment. It can be further seen in Fig. 4b that the region of highest temperature (742 K) is on the ceiling, directly above the fire, while the temperature reduces as the distance from the source/fire increases. As there is a gap between the ceiling of the compartment and the evacuation doors, the hot air is trapped on the upper portions of the compartment, while the lower sections of the compartment are at comparatively lower temperature.

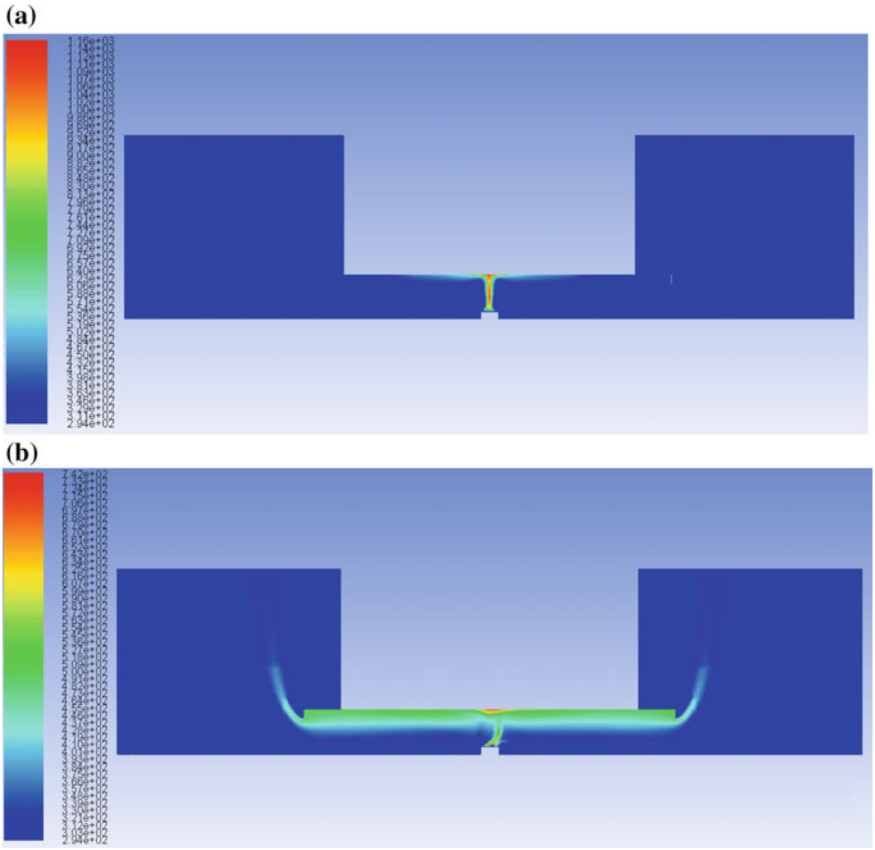


Fig. 4 Static temperature variations after **a** 8 s and **b** 360 s of fire eruption

The evolution of fire can be clearly seen in the figures. Only a small region of the compartment is affected 8 s after the fire has erupted. However, as the time increases, most of the compartment is filled with smoke, and hence most of the volume of the railway compartment is occupied by higher temperature zones.

Data is omnipresent. Data on its own is useless unless it is intelligently analysed and understood. It has been reported that the reliability of many research investigations carried out by many reputable organizations is coming under increasing scrutiny. It is also true that the entrenched culture of cut-throat competition and fraudulent behaviour is hindering the progress. All reliability investigations are heavily dependent upon the quality of data and the intelligent extracting capability of individuals. Extracting the intelligence from Big and Open Data is a challenging task indeed.

For further analysing the flow behaviour within the railway compartment, profiles of static temperature have been drawn on (a) a vertical line directly above the methane burner (Fig. 5a), and (b) a horizontal line in the middle of the compartment

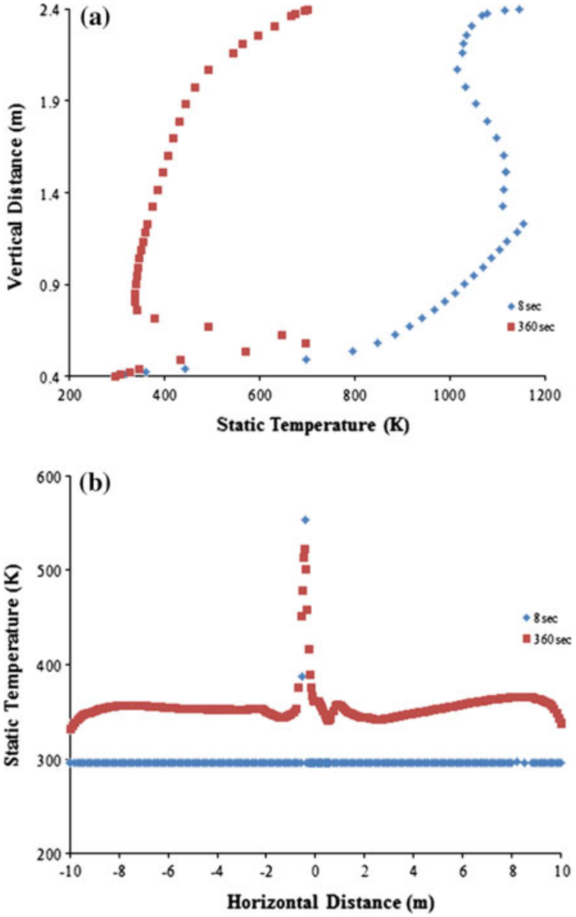


Fig. 5 Static temperature profiles **a** above the methane burner and **b** along the length of the compartment

from one evaluation door to the other (Fig. 5b). Figure 5a depicts that just above the methane burner, where the fire has erupted, the static temperature rises considerably; however, just after the fire, it drops back and then increases gradually while going towards the ceiling of the compartment. Furthermore, it can be seen that the fire is still pre-mature after 8 s of eruption.

Figure 5b depicts that after 8 s of fire eruption, only the region directly above the fire is at higher temperature, whereas the rest of the compartment is at ambient temperature. However, after 360 s, although the same trend follows but the ambient temperature within the compartment has increased significantly.

Further analysing the flow behaviour within the compartment, Figs. 6a, b depict the variations in the flow velocity magnitude after 8 and 360 s respectively. It can be seen that as the fire erupts, due to the higher temperature directly above the fire, the flow velocity increases, and then spreads outwards. This trend is similar to the one observed in case of temperature variations within the compartment at the same occasions. However, it should be noted that once enough time has elapsed, the flow

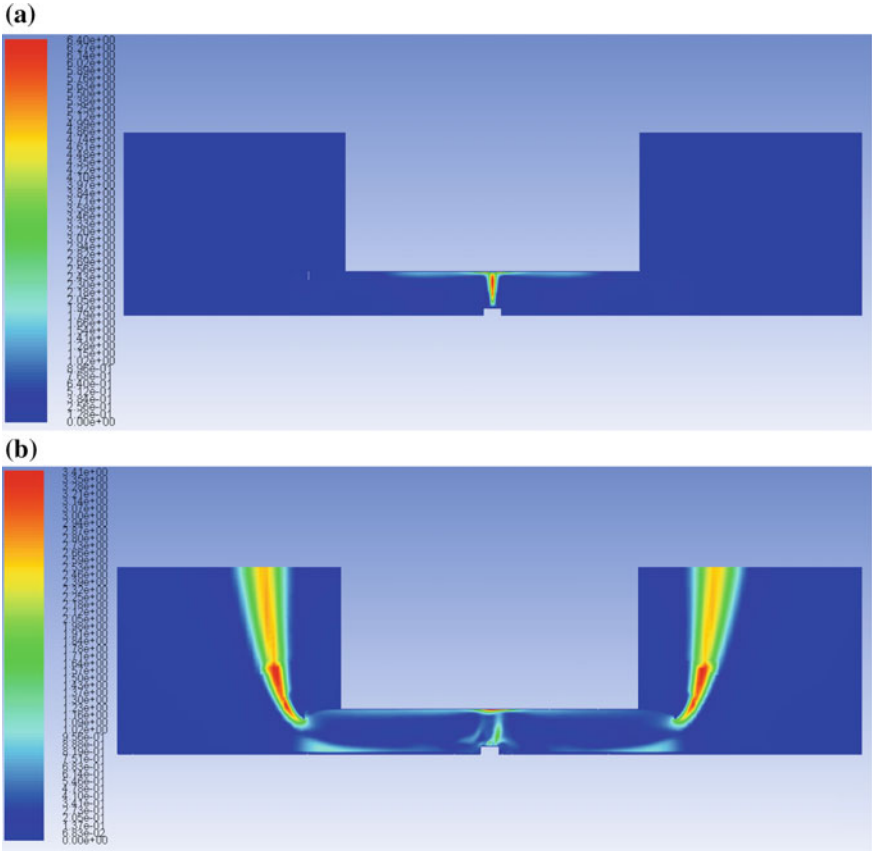


Fig. 6 Velocity magnitude variations after **a** 8 s and **b** 360 s of fire eruption

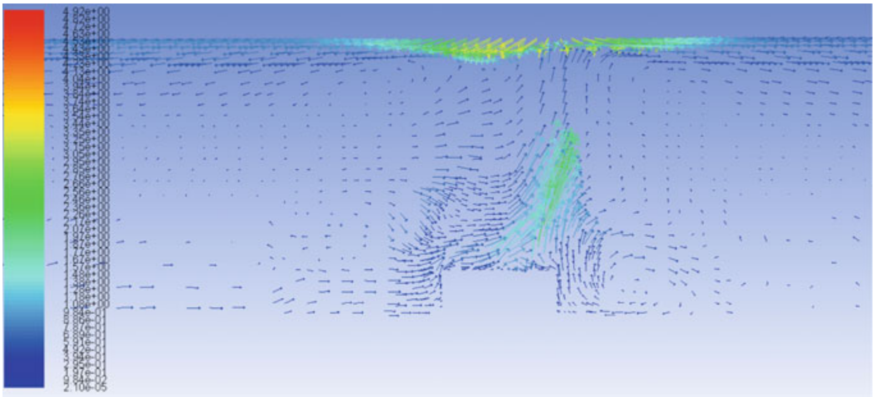


Fig. 7 Velocity vectors in the vicinity of the methane burner after 360 s of fire eruption

velocity within the compartment reduces significantly, while the smoke is rushing out of the evacuation doors at significantly higher velocities.

Figure 7 depicts the velocity vectors in the vicinity of the methane burner. It can be seen that the smoke from the fire travels upwards and then sideways to the evacuation doors, while, at the same time, fresh air is entering the compartment from the lower sections. It can be further noticed that the region in the centre of the compartment is almost stationary.

Figure 8 depicts the variations in the flow velocity magnitude at different time instants both vertically above the methane burner, and horizontally along the centre of the compartment. Figure 8a depicts that, after 8 s of fire eruption, just above the methane burner, the flow velocity increases significantly, whereas the flow is stationary at the ceiling due to no-slip boundary condition. Furthermore, after 360 s of fire eruption, the flow velocity just above the methane burner is considerably higher. The flow than slows down until it reaches a certain height from where onwards, it starts spreading, and hence the flow velocity increases.

Figure 8b depicts that only the region directly above the fire is at higher velocity. However, there is an indication that the flow again starts to accelerate near the evacuation doors. This has already been observed in Fig. 6b.

The temporal distribution of the molar concentration of CO_2 within the railway compartment is depicted in Fig. 9, where Fig. 9a represents the scenario after 8 s of fire eruption, while Fig. 9b represents the scenario after 360 s of fire eruption within the compartment. It can be seen that as the fire erupts within the railway compartment, the smoke is generated, which, due to being lighter than air, travels against the gravitational force, until it comes in contact with the ceiling of the compartment. Then it starts to spread outwards in both directions until it escapes out from the evacuation doors, and into the environment. It can be further seen in Fig. 9b that the smoke is trapped on the upper portions of the compartment, where its concentration is almost uniform, while the lower sections of the compartment are free of smoke. The evolution of smoke can be clearly seen in the figures. Only a

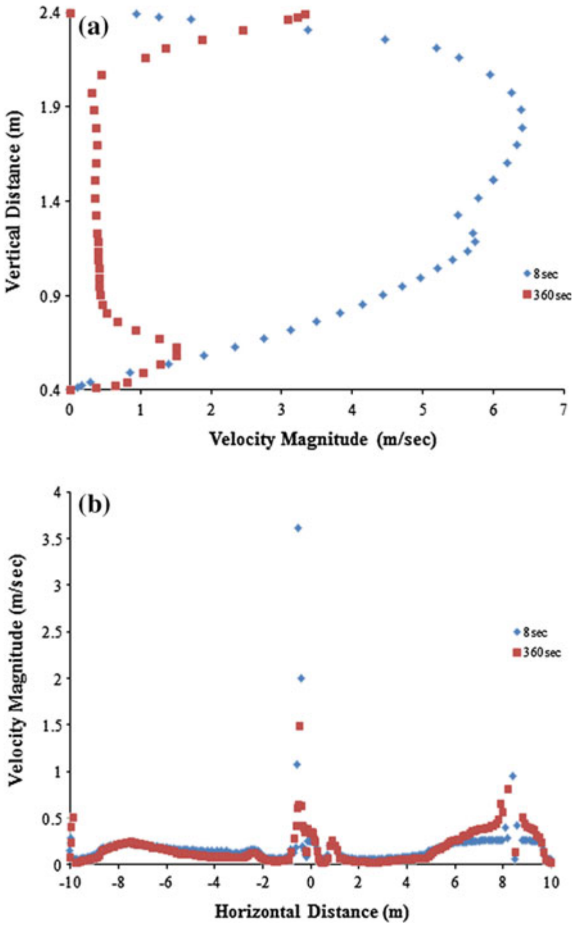


Fig. 8 Velocity magnitude profiles **a** above the methane burner and **b** along the length of the compartment

small region of the compartment is filled with smoke 8 s after the fire has erupted. However, as the time increases, most of the compartment is filled with smoke.

Figure 10 depicts the molar concentration of CO_2 directly above the methane burner, both after 8 and 360 s of fire eruption. It can be seen that in the early stages of fire, as it is still pre-mature, a lot of CO_2 is being ejected into the compartment, although the overall region of the compartment occupied by the smoke is relatively small. However, after 360 s, the CO_2 molar concentration is highest near the methane burner, which then increases gradually towards the ceiling of the compartment. CO_2 molar concentration in the centre of the compartment, from one door to the other, has not been shown in the present study as this region is almost free of CO_2 for the range of parameters considered in the present study. Hence, further studies need to be conducted in order to widen the range of CO_2 concentration analysis.

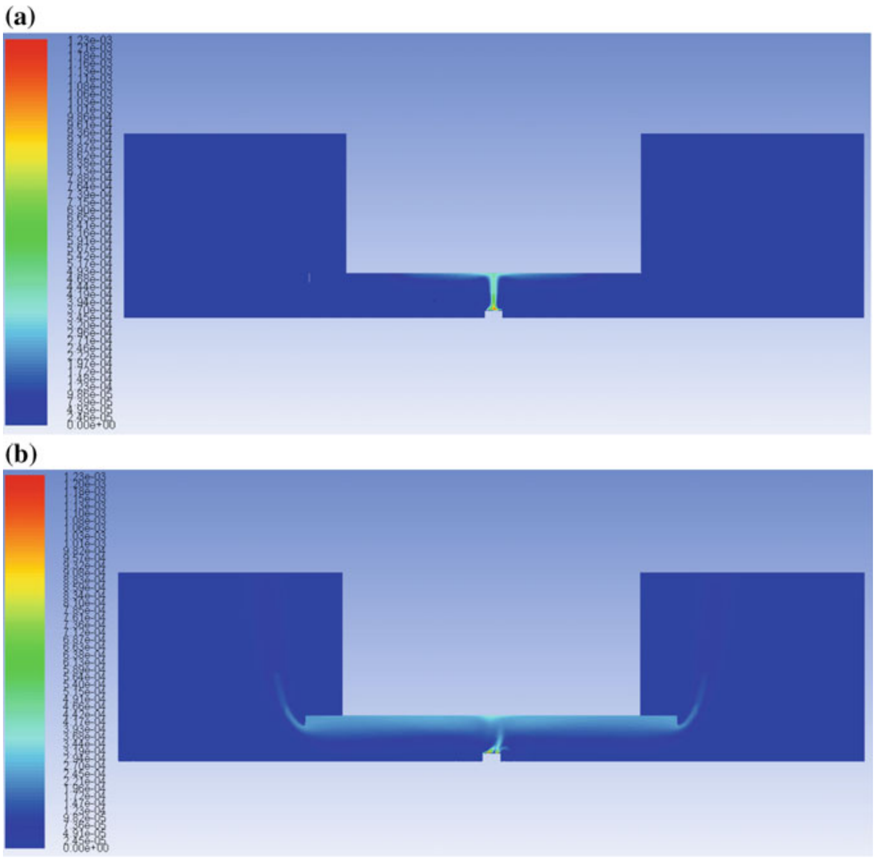
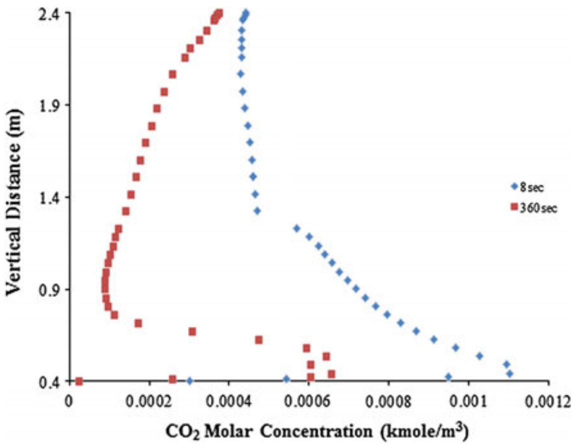


Fig. 9 CO_2 molar concentration variations after **a** 8 s and **b** 360 s of fire eruption

Fig. 10 CO_2 molar concentration profiles above the methane burner



4 Conclusion

Unavailability, poor quality, unreliable assets drive nations' invaluable resources to unsustainable and unrecoverable bottomless pit of misery and poverty of unimaginable dimension. Our 'Quality of life' and 'happiness' is heavily dependent upon reliable and sustainable performance of all assets under all operational conditions. Best practices, best guidelines and national/international standards are the only way to reduce uncertainties and enhance reliability of our assets through smart management practices.

A detailed CFD based investigations on a railway compartment fire has been carried out in the present study. Three primary parameters i.e. the static temperature, the flow velocity and the molar concentration of CO_2, have been numerically analysed within a railway compartment. The spatio-temporal variations of these parameters indicate that as a fire erupts in a railway compartment, the smoke at higher temperature rises up and comes in contact with the ceiling of the compartment. Then it spreads outwards towards the evacuation doors on either ends of the compartment. The upper region of the compartment is filled with smoke, containing a large amount of CO_2, while the lower section of the compartment is relatively free of smoke and CO_2. It has further been noticed that the upper section of the railway compartment is at a higher temperature as compared to the lower section, and similarly the flow velocity is significantly higher in the top section of the compartment. Moreover, it has been observed that the smoke exits the compartment through the upper part of the evacuation doors, while fresh air enters the compartment from the lower part of these doors. Hence, CFD can be used as an effective tool in order to analyse railway compartment fires.

References

1. Mo S, Li Z, Liang D, Li J, Zhou N (2013) Analysis of smoke hazard in train compartment fire accidents base on FDS. Proc Eng 52:284–289
2. White N (2010) Fire development in passenger trains. Master thesis. Centre for environment safety and risk engineering. Victoria University, Australia
3. Dowling V, White N (2004) Fire sizes in railway passenger saloons. Fire Saf Sci 6: pp 6b-3–1 http://www.iafss.org/publications/aofst/6/6b-3 Accessed 30 Nov 2014
4. Chiam BH (2005) Numerical simulation of a metro train fire, fire engineering research report 05/1. Department of Civil Engineering. University of Canterbury, New Zealand
5. Babrauskas V, Peacock RD (1992) Heat release rate: the single most important variable in fire hazard. Fire Saf J 18(3):255–272
6. Jahn W, Rein G, Torero JL (2009) The effect of model parameters on the simulation of fire dynamics. Fire Saf Sci 9:1341–1352
7. Rusch D, Blum L, Moser A, Roesgen T (2008) Turbulence model validation for fire simulation by CFD and experimental investigation of a hot jet in crossflow. Fire Saf J 43 (6):429–441

8. Andreini A, Da Soghe R, Giusti A, Caruso L (2011) Pyrolysis modeling and numerical simulation of rail carriage fire scenarios for the safe design of a passenger train. Department of Energy Engineering. University of Florence, Italy
9. Mishra R, Singh SN, Seshadri V (1998) Study of wear characteristics and solid distribution in constant area and erosion-resistant long-radius pipe bends for the flow of multisized particulate slurries. Weir 217(2):297–306
10. Binbin W (2011) Comparative research on FLUENT and FDS's numerical simulation of smoke spread in subway platform fire. Proc Eng 26:1065–1075
11. ANSYS® Academic Research, Release 15.0, Help System, Theory Guide. ANSYS, Inc
12. Tesfa B, Mishra R, Gu F, Ball A (2011) Combustion characteristics of CI engine running with biodiesel blends. International conference on renewable energies and power quality, Las Palmas de Gran Canaria, Spain

Safety Analysis of Mining Machines Specific Maintenance Operations

Ljubisa Papic, Srdja Kovacevic, Diego Galar and Adithya Thaduri

Abstract By the rule, the object of the safety analysis is the technical system, for example, production or transportation, as the mining machines or their technological equipment are. The maintenance operations, up to present days, haven't been investigated as the subject of safety analysis. However, as the practice of the technical systems maintenance shows so far many maintenance operations contain causes of danger. It means that it is useful to analyze such operations from the safety standpoint. From the standpoint of the mining machines safety, it should be stressed that in some researches, the expression "specific" is used for the critical maintenance operations. Therefore, the safety analysis of maintenance operations should precede the stage of maintenance operations. The possible approaches to the safety analysis in the area of maintenance on the basis of the method Failure Modes, Effects and Criticality Analysis are presented in the paper.

Keywords Safety analysis · Critical maintenance · Operations

1 Introduction

The most important characteristic of quality of the technical systems maintenance is the safety that, according to the Dictionary of technical regulations [1], means the absence of proscribed (unacceptable, intolerable) risk. Therefore, the safety analysis

L. Papic (✉)
University of Kragujevac, Kragujevac, Serbia
e-mail: dqmcenter@mts.rs

S. Kovacevic
JP PK Kosovo Obilic, Belgrade, Serbia
e-mail: srdja.kovacevic@mts.rs

D. Galar · A. Thaduri
Luleå University of Technology, Luleå, Sweden
e-mail: diego.galar@ltu.se

A. Thaduri
e-mail: adithya.thaduri@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_35

of maintenance operations should precede the stage of maintenance operations. Several methods for the investigation of technical systems safety were proposed in the design phase [2, 3]. One of the most complete methods for the safety analysis is given in the paper [4].

2 Maintenance Operations of Mining Machines Safety Analysis

By the rule, the object of the safety analysis in the technical system, for example, production or transportation, as the mining machines or their technological equipment is important. The maintenance operations, up to present days, haven't been investigated as the subject of safety analysis. However, as the practice of the technical systems maintenance shows so far many maintenance operations contains causes of danger. It means that it is useful to analyze such operations from the safety standpoint. From the standpoint of the mining machines safety, it should be stressed that in some researches, with the results presented in the references [5, 6], the expression "specific" is used for the critical maintenance operations. The possible approaches to the safety analysis in the area of maintenance on the basis of the method Failure Modes, Effects and Criticality Analysis (FMECA) [7], are presented further on in the text. The objective of FMECA method is contained in the analysis of the effects of maintenance operations on the safety of the technological operational process of the mining machines [8].

The same operation of the mining machine maintenance, depending on the situation, can be analyzed from the standpoint of estimation of its effect on the safety of the manufactured product (for example, repaired or revitalized mining machine) or on the safety of a process of its production (for example, technological process of the mining machine maintenance) [9, 10]. For example, the welding operation, executed in the procedure of mining machine assemblage (bucket wheel excavator, landfill machine, dumping machine, self/transporter bandwagon, dragline dredger) after a heavy accident, can be evaluated from the standpoint of its effect on the safety of the machine or on the assemblage process [11, 12]. Therefore, each maintenance operation, from the standpoint of safety, is useful to analyze in two aspects:

- as the result of performing the maintenance operation that can be presented in the form of the overhauled mining machine [13],
- as the process when the maintenance operation is performed, which can be shown in the form of technological process of the mining machine maintenance [14].

In that way, depending on the danger source, safety of the maintenance operation can be conditioned by [15]:

- the result of the maintenance operation performing,
- the procedure of maintenance operation realization,
- the result of the maintenance operation performing and procedure of maintenance operation realization.

3 Investigation of Dependence Between Precision and Capability of Mining Machines Maintenance Operations

Since the result of performing the maintenance operation represents an overhauled mining machine, it is useful to establish the degree of criticality of the maintenance operation/operations. The idea to have the analysis of the maintenance operation/operations criticality consists of the following factors [16]:

- frequency of the failure modes (frequency of defect mode), caused by the loss of maintenance operation precision,
- probability of failure mode detection,
- consequences caused by the failure mode (defect mode).

Here the precision of maintenance operation/maintenance technological process means its characteristic to ensure the closeness of real and nominal parameters of the mining machines that have been maintained (product that has been manufactured) [16]. The disturbance in the maintenance operation precision leads to the failure mode of mining machine (to the defect of the product that has been manufactured) [17]. The parameter natural zone of dispersion (quality characteristic) is compared with the zone of specification limits, in order to analyze the maintenance operation precision. Natural zone of dispersion is the area of parameter values of a mining machine that has been maintained (product that has been manufactured), with the precision probability close to 1. At the normal distribution of the parameter of the mining machine that has been maintained (product that has been manufactured), the natural zone of dispersion is assumed to be $\pm 3\sigma$, i.e. 6σ, where σ—is standard deviation of the mining machine (product that has been manufactured) parameter. In that case, the natural zone of dispersion is determined as the zone of parameter values that corresponds to the probability of its precision of 0.9973. The natural zone of dispersion or real dispersion, as the density function of normal parameter distribution f(x), is presented on Fig. 1.

The following main indicator—process capability factor Cp that signifies the ratio of zone of specification limits T and natural zone of dispersion of the parameter that is analyzed [18], can be used to analyze the precision of operation for maintaining the mining machines (technological process):

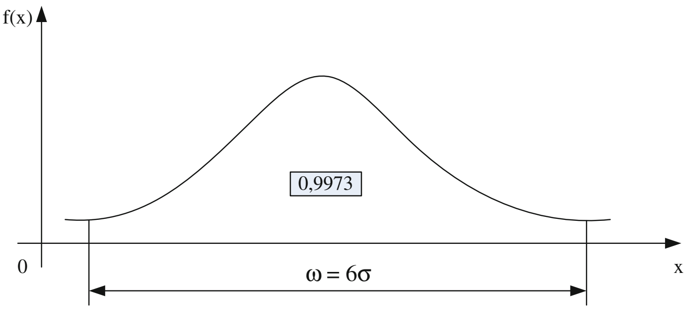


Fig. 1 Graphical presentation of natural zone of dispersion in the case of density function of normal parameter distribution

$$Cp = T/\omega = T/6\sigma. \tag{1}$$

It should be noted that in the relevant literature, dedicated to the research of the technological processes [19], the introduced ratio is named precision coefficient.

The lager value of Cp, less the defect level (smaller number of failure modes) δ, which is provided by certain (specific) maintenance operation. There is a singular dependence between Cp value and the level of defects (number of failure modes) δ (in the case of normal parameter distribution of a mining machine that is maintained—product that is manufactured). This dependency is shown in Table 1.

For the parameter of a mining machine that is maintained, and which distribution differs from normal, the precision of the maintenance operation is characterized by the precision coefficient Pc which is determined by the expression:

$$Pc = T/2.745\mu, (Pc = 1/Cp), \tag{2}$$

where it is assumed that he distribution of the parameter is close to Rayleigh distribution [19], where μ—is medial parameter value.

This expression should be used for the parameters of mining machine that s maintained, and that is characterized by:

- non-coaxiallity of two nominally axial cylindrical surfaces (eccentricity),
- non-parallel pair of surfaces,
- non-normality pair of surfaces or axe and surface,
- difference of walls.

Table 1 Relation of process capability factor and level of defects (number of failure modes)

C_p	2	1.67	1.33	1.00	0.83	0.71	0.63	0.56
δ %	2 PPB	6 PPM	63 PPM	0.27	1.30	3.30	4.90	9.30

Note
1 PPB—1 defect (failure mode) on a billion overhauled mining machines (units)
1 PPM—1 defect (failure mode) on a million overhauled mining machines (units)

Table 2 Relation between precision coefficient and level of defects (number of failure modes)

P_c	1.10	1.00	0.90	0.80	0.70	0.60
δ [%]	0.01	0.20	0.90	2.00	5.10	12.00

Interdependence of precision coefficient Po and expected level of defects (number of failure modes) δ is presented in Table 2.

4 Selection of the Most Important Maintenance Operations of Mining Machines

Once we clarified this relation, we can go to the calculation of the criticality of maintenance operation that is performed in order to analyze the effect of (technological) maintenance operation on the safety of a mining machine that is maintained. The schematic that considers factors in the calculation of criticality degree of a mining machine maintenance operation, is shown on Fig. 2.

Criticality degree of a mining machine maintenance operation is calculated by:

$$\mathrm{Ci} = \mathrm{B1i} \cdot \mathrm{B2i} \cdot \mathrm{B3i} \tag{3}$$

where:

B1i estimation of frequency (probability) of potential defect occurrence (failure mode) of i-th element (or item) of mining machine,

B2i estimation of probability of defect (failure mode) detection of i-th element (or item) of a mining machine, before it is delivered to the user,

B3i estimation of seriousness of defect (failure mode) consequences of i-th element (or item) of a mining machine

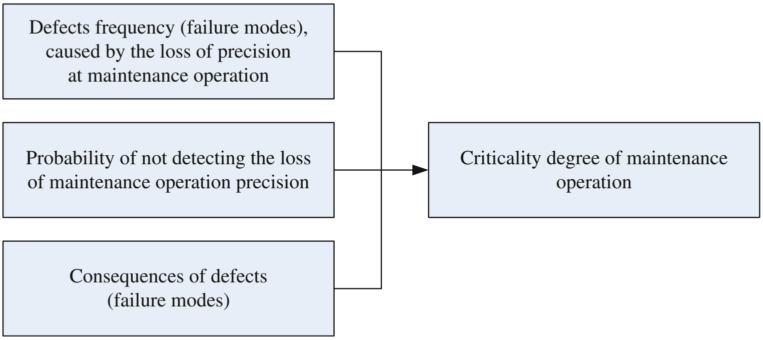

Defects frequency (failure modes), caused by the loss of precision at maintenance operation
Probability of not detecting the loss of maintenance operation precision
Consequences of defects (failure modes)
Criticality degree of maintenance operation

Fig. 2 Calculation of criticality degree of maintenance operation

Criticality degree of defects (failure modes) of a mining machine is calculated by:

$$\mathrm{C} = \sum_i C_i \tag{4}$$

In that way, if the source of danger caused by an overhauled mining machine, the safety analysis regarding FMECA—object is analogous to the safety analysis of the principally new mining machine of the same usage. Regarding the previous, the detailed safety analysis of the overhauled mining machine applying FMECA method would not be considered.

Another source of danger is a technological maintenance process. In that case, FMECA—process can be applied for the safety evaluation, that analysis the influence of process operation on the safety of:

- mining machine submitted to the maintenance,
- technological maintenance process.

Technological maintenance process has to be divided to maintenance operations. One operation, depending on the situation, can be analyzed from the standpoint of its influence on the safety of mining machine submitted to the maintenance tasks or from the standpoint of evaluation of its influence on the safety of working process (maintenance process) [20].

In that way, the criticality degree of an operation is calculated by the formula analogous to formula (3). Therefore, value of the coefficient B1 can be found using Table 3, considering process capability factor Cp (or precision coefficient Pc). It should be noted that value Cp (or Pc) can be evaluated with an expertise by a working group (working team) or calculated by a selection of random samples with a later analysis of precision of technological maintenance process [21].

Table 3 Values of coefficient B1 (FMECA of technological maintenance process)

Description of defect frequency (failure mode)	Adjoined value of C_p	Value of B_1 (points)
Defect (failure mode) practically not possible	<1.67	1
Very rare occurrence of defect (failure mode)	1.33	2
Rare occurrence of defect (failure mode)	1.00	3
Defect (failure type) is possible	0.83	4
Defect (failure mode) is highly possible	0.71	5
Defect (failure mode) occurs frequently	0.63	7–8
Very frequent occurrence of defect (failure mode)	>0.56	9–10

Table 4 Values of coefficient B2 (FMECA of technological maintenance process)

Description of probability to detect precision disruption	Value of B_2 (points)
Very high probability of precision loss detection, since the event is easily identified (recognized)	1–2
High probability of precision loss detection	3–4
Moderate (medium) probability of precision loss since the event is hard to identify (recognize)	5–6
Low probability of precision loss detection	7–8
Very low probability of precision loss detection. The event can not be identified (recognized)	9–10

Table 5 Values of coefficient B3 (FMECA of technological maintenance process)

Consequences of defect (failure modes)	Value of B_3 (points)
Not significant. Purchaser (user) can notice them	1–2
Significant. Maintenance of object (technical system, mining machine) can be performed at the purchaser (user) site with insignificant cost	3–4
Very significant. Maintenance cost are significant and caused by object (technical system, mining machine) stoppage	5–6
Critical. Defect (failure mode) generates loss (accident) of object (technical system, mining machine). There is no danger for the people and environment safety	7–8
Vary critical. Defect (failure mode) is connected with people and environment safety	9–10

Value of coefficient B2, which characterizes a probability of detecting the facts of precision disruption, can be found using Table 4, depending on usual (accepted) way of control of a technological maintenance process.

Value of coefficient B3, which characterizes the consequences of a defect (failure modes), and which occurs as the result of precision loss, can be determined using Table 5.

Beside the evaluation of effect of a technological maintenance process on mining machine precision that is maintained, method of FMECA—process enables the influence of maintenance operation on the technological maintenance process to be evaluated [22]. The influence of a maintenance operation on the safety of the technological maintenance process shows that failure in operation, i.e. disturbance of one or several characteristics of the maintenance operations, can cause the critical consequences for the maintenances process (influence on maintenance object—mining machine—is not taken in account).

Analysis of maintenance operation criticality starts from the division of a technological process to the individual operations. Further on, the analysis of possible dangers as the result of potential disruptions of the maintenance operation

Table 6 Format of a table for analyzing dangerous situations

Operation	Description of dangerous situation	Cause of operation disruption	Dangerous event	Parameters of condition
–	–	–	–	–

Table 7 Values of coefficient B1 (FMECA of maintenance operation)

Description of disruption frequency of maintenance operation	Values of B_1 (points)
Disruption of maintenance operation practiclly is not possible	1
Disruption frequency of maintenance operation is small	2–5
Medium disruption frequency of maintenance operation	6
High disruption frequency of maintennce operation	7–8
Very high frequency, disruption of maintenance operation is certain to occur	9–10

Table 8 Values of coefficient B2 (FMECA of maintenance operation)

Description of detection probability of maintenance operation disruption	Values of B_2 (points)
Very high detection probability of maintenance operation disruption, since the event is easy to identify (recognize)	1–2
High detection probability of maintenance operation disruption, identifiction (recognition) of the event is simple	3–4
Moderate (medium) detection probability of maintenance operation disruption, the event is difficult (complicated) to identify (recognize)	5–6
Low detection probability of maintenance operation disruption	7–8
Very low detection probability of maintenance operation disruption. This event can not be idenfied (recognized)	9–10

is performed. The results of this analysis are useful to organize in the table form, as shown in Table 6.

Further calculation of criticality degree is performed for the maintenance operations that are recognized (evaluated, considered to be) as the most significant for the safety of a technological maintenance process. The calculation of criticality degree of a maintenance operation, regarding safety of the technological maintenance process can be executed by a formula analogous to formula (3). Coefficient B1 is selected from Table 7.

Coefficient B2, which characterizes the possibility of disruption detection of maintenance operation, is selected using Table 8. The consequences of disruption of maintenance operation are described using coefficient B3, with the values determined using Table 9. Separation of the most significant maintenance operations is performed by comparing the criticality degree of i-th maintenance operation Ci with the limit value of Ccrit = 125. If Ci > Ccrit, the i-th maintenance operation is considered to be critical, so that the use of correction measures is compulsory in the

Table 9 Values of coefficient B3 (FMECA of maintenance operation)

Consequences of disruption of maintenance opeation	Values of B_3 (points)
Insignificant. Disruption of maintenance operation is easily removed	1–2
Significant. Disruption of maintenance operation leads to equipmet stoppage and disrupts technolgical maintenance process	3–4
Very significant. Disruption of maintenance operation causes the performing of maintenance tasks to stop	5–6
Critical. Disruption of maintenance operation causes the execution of maintenance tasks to stop and can cause accident. There is no danger for the safety of people and environment	7–8
Very critical. Disruption of maintenance operation is connected to the safety of people and environment	9–10

framework of quality management system, according to the requirements of international standard ISO 9001, in order to reduce the criticality degree of i-th maintenance operation.

5 Qualitative Analysis of Limiting Parameters of Technological Process of Mining Machines Maintenance

The selection of the main (limiting) parameters of technological maintenance processes that limit the safety of the mining machines is performed in order to reach the following [23]:

- correction of the way of quality control at performing the maintenance operations,
- conducting the measures for improving (rationalization) technological maintenance process,
- specification (precise determination) of the demands for purchasing the assemblies and material.

The selection of the limiting parameters of technological maintenance process is performed on the basis of the analysis of failure causes of mining machines according to the order shown on Fig. 3. As the result of the analysis of failure

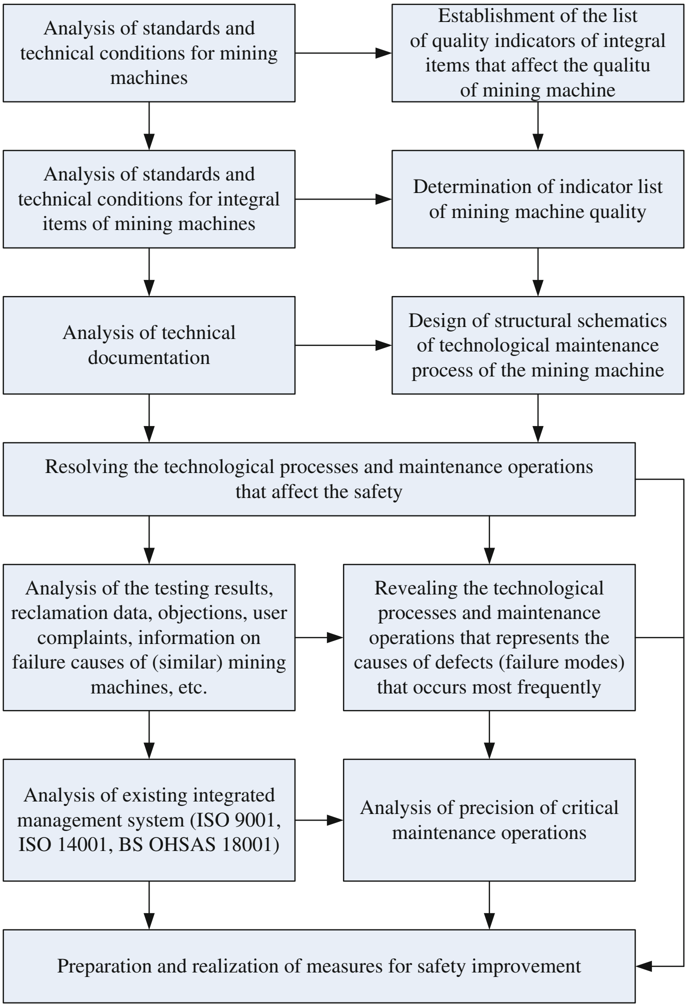

Analysis of standards and technical conditions for mining machines
Establishment of the list of quality indicators of integral items that affect the qualitu of mining machine
Analysis of standards and technical conditions for integral items of mining machines
Determination of indicator list of mining machine quality
Analysis of technical documentation
Design of structural schematics of technological maintenance process of the mining machine
Resolving the technological processes and maintenance operations that affect the safety
Analysis of the testing results, reclamation data, objections, user complaints, information on failure causes of (similar) mining machines, etc.
Revealing the technological processes and maintenance operations that represents the causes of defects (failure modes) that occurs most frequently
Analysis of existing integrated management system (ISO 9001, ISO 14001, BS OHSAS 18001)
Analysis of precision of critical maintenance operations
Preparation and realization of measures for safety improvement

Fig. 3 Scheme of parameter analysis of technological maintenance processes of mining machines that influences the safety

causes of mining machines, certain output parameters of the technological maintenance process should be determined previously, which disruption may potentially lead to the loss of operational capability of the mining machines.

6 Conclusion

The proposed approach enables the safety of technological process and (specific) operations of mining machines maintenance to be analysed in the same framework by using the FMECA method. It is useful to supplement the analyses of types, consequences and criticality of failure modes of technological process of mining machines maintenance by qualitative analysis in order to establish the main parameters of the maintenance process that restrictively acts on the safety of the mining machines. The main parameters of the technological process of mining machines maintenance are parameters of process involving spare parts and assemblies manufacture on which the exploitation characteristics substantially depend on, including safety of the mining machines.

References

1. Аронов ИЗ, Теркель АЛ, Рыбакова АМ (2006) Словарь – справочник по техническому регулированию, Стандарты и качество, Москва, 288 с
2. Dhillon BS (2003) Engineering safety, fundamentals, techniques and applications, World Scientific, New Yersey, p 239
3. Zio E (2009) Computational methods for reliability and risk analysis. World Scientific, New Jersey, p 362
4. Александровская ЛН, Афанасьев АП, Лисов АА (2008) Современные методы обеспечения безотказности сложных технических систем, Логос, Москва, 208 с
5. Pantelić M (2011) Tehnologije održavanja rudarskih mašina na površinskim ugljenokopima, Janko Stajčić, Lazarevac, 522 s
6. Pantelić M (2009) Unapređenje koncepcije održavanja putem operativnog upravljanja sigurnošću bagerskih jedinica na površinskim kopovima, Doktorska disertacija, Tehnički fakultet, Čačak, 243 s
7. Stamatis DH (2003) Failure mode effect analysis: FMEA from theory to execution, 2nd edn. American Society for Quality, Milwaukee, p 487
8. Sahoo T, Sarkar PK, Sarkar AK (2014) Maintenance optimization for critical equipments in process industries based on FMECA Method. Int J Eng Innov Technol 3(10)
9. Barabady J, Kumar U (2008) Reliability analysis of mining equipment: a case study of a crushing plant at Jajarm Bauxite mine in Iran. Reliab Eng Syst Saf 93(4):647–653
10. Gustafson A, Schunnesson H, Galar D (2011) Maintenance indicators for underground mining equipment a case study of automatically versus manually operated LHD machines. In: Proceedings of the 24th international congress on condition monitoring and diagnosis engineering management: COMADEM 2011. COMADEM International Stavanger
11. Gustafson A, Schunnesson H, Galar D, Kumar U (2013) The influence of the operating environment on manual and automated load-haul-dump machines: a fault tree analysis. Int J Min Reclam Environ 27(2):75–87
12. Kumar U, Klefsjö B, Granholm S (1989) Reliability investigation for a fleet of load haul dump machines in a Swedish mine. Reliab Eng Sys Saf 26(4):341–361
13. Jardine AK, Tsang AH (2013).Maintenance, replacement, and reliability: theory and applications. CRC press, Boca Raton
14. Liyanage JP, Kumar U (2003) Towards a value-based view on operations and maintenance performance management. J Qual Maint Eng 9(4):333–350

15. Laurence D (2005) Safety rules and regulations on mine sites—the problem and a solution. J Saf Res 36(1):39–50
16. Суливанов МН, Фридман АЭ, Кудряшова ЖФ (1987) Качество измерений, Лениздат, Ленинград, 295 с
17. Markeset T, Kumar U (2003) Design and development of product support and maintenance concepts for industrial systems. J Qual Maint Eng 9(4):376–392
18. Papić Lj (2011) Menadžment kvalitetom, DQM, Prijevor, 288 s
19. Крюков СП, Бордунов СД, Александровская ЛН, Аронов ИЗ, Захаревич АП, Кузнецов АГ, Кушельман ВЯ (2007) Методы анализа и оценивания рисков в задачах менеджмента безопасности сложных технических систем, Аэрокосмическое оборудование, Санкт Петербург, 460 с
20. Parida A, Kumar U (2006) Maintenance performance measurement (MPM): issues and challenges. J Qual Maint Eng 12(3):239–251
21. Hall R (1997) Analysis of mobile equipment maintenance data in an underground mine. Unpublished Master's Thesis, Queen's University, Kingston, Ontario
22. Bevilacqua M, Braglia M (2000) The analytic hierarchy process applied to maintenance strategy selection. Reliab Eng Sys Saf 70(1):71–83
23. Novak T, Kohler JL (1998) Technological innovations in deep coal mine power systems. IEEE Trans Ind Appl 34(1):196–204

Experiences and Insights in Development of Probabilistic Safety Assessment of Research Reactors in BARC

Sachin Kumar, N.S. Joshi, Vivek Mishra and P.V. Varde

Abstract Probabilistic Safety Assessment (PSA) techniques have become a standard tool in safety evaluation of nuclear power plant. For research reactors, experiment facilities aspects add a further dimension to application of PSA. Such experiments may increase the risk of damaging the reactor core. Sufficient experience is now available to suggest that current PSA methodology is a valuable tool for application to research reactor. Level 1 PSA of three research reactors i.e. 40 MWt Cirus, 100 MWt Dhruva, 2 MWt Upgraded Apsara Reactor Project in BARC, Trombay has been performed. The modelling for PSA of these research reactors has been carried out considering only with the internal initiating events for full power operational state and reactor core as the source of radioactive inventory. The special issues like common cause failure, uncertainty analysis and human reliability analysis have been addressed in the analysis. The specific purpose of these studies apart from normal application, i.e. to obtain safety insights, has been (a) design evaluation, (b) evaluation of scenario involving multiple failures, (c) regulatory support in decision making, (d) ageing assessment and (e) development of future applications like risk-monitor, risk-based ISI programme, and surveillance test interval optimization for safety systems. Further work is in progress on external event PSA e.g. Seismic PSA, and Aircraft Impact analysis and Fire PSA methodology. This paper discusses the experiences and insights in the development of PSA of Research Reactors at Bhabha Atomic Research Centre, Trombay.

Keywords Probabilistic Safety Assessment · Core Damage Frequency · Initiating events · Accident sequence analysis · Risk monitor

S. Kumar (✉) · N.S. Joshi · V. Mishra · P.V. Varde
RRSD, BARC, Mumbai 400 085, India
e-mail: s_kumar@barc.gov.in

N.S. Joshi
e-mail: nsjoshi@barc.gov.in

V. Mishra
e-mail: vmishra@barc.gov.in

P.V. Varde
e-mail: varde@barc.gov.in

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_36

1 Introduction

Probabilistic Safety assessment (PSA), also known as Probabilistic Risk Assessment (PRA), of nuclear reactors, essentially aims at identifying the events and their combinations that can lead to severe accidents, assessing the probability of occurrence of each combination and evaluating the consequence. Since the completion of the landmark Reactor Safety Study (commonly referred to as WASH-1400) [1] in 1975, PSA results and insights have been used to support regulatory decision making regarding nuclear power plants (NPPs). The assurance of the safe operation of NPPs around the world and the prevention of incidents in these installations remains a key concern of the nuclear community. PSA has been used in nuclear installations for assessing the safety of the components and the plant. The PSA technique complements the conventional deterministic methodology in safety assessments of design basis scenarios. At the same time, the insight obtained using PSA methods on 'beyond design basis' scenarios supplement the deterministic findings. It is desirable to perform all three levels of PSA for a nuclear reactor. As a minimum requirement, plant should carry out Level 1 PSA with internal and external events as applicable to the plant [2]. It is important to recognize that the experimental facilities in the research reactors add a further dimension to the application of PSA. Such experiments might not only contribute directly to radioactive release to the atmosphere but also in some cases have potential to increase the risk of damaging the reactor core. The numbers of IAEA-TECDOCs available for research reactor PSA shows the interest of international community in embracing this technique for integrated safety assessment of the research reactors.

In Bhabha Atomic Research Centre (BARC) at Trombay, power level of nuclear research reactors varies from few watt levels to hundreds of MW level. In BARC, Level 1 + PSA of the operating research reactor Dhruva had been carried out in the year 2002 [3]. Level 1 PSA of research reactor Cirus, which was permanently shutdown in the year 2010, was completed in the year 2009 [4]. Level-1 PSA of Upgraded Apsara Research Reactor Project was completed in the year 2012 [5]. The work carried out in these studies involve selection of initiating events, event tree modeling, data collection and analysis, system modeling, accident sequence modeling and finally estimation of core damage frequency (CDF). In these studies, modeling has been carried out considering only the internal initiating events for full power operational states, and the reactor core as the source of radioactivity. The special issues like common cause failure, uncertainty analysis and human reliability analysis have been addressed in the analysis.

2 Level 1 + PSA of Dhruva Reactor

Dhruva reactor, 100 MWt, high neutron flux (maximum $\sim 1.8 \times 10^{14}$ n/cm^2/sec) research reactor located at BARC, has the distinction of being one of the high flux research reactors in the world. Dhruva was commissioned in 1985 and achieved full power operation in 1988. In this reactor, natural uranium is used as fuel and heavy water as coolant, moderator and reflector.

In the year 1997, a program for a detailed Level 1 PSA of Dhruva research reactor was initiated [3]. The PSA code PSAPACK version 4.2, supplied by IAEA was used as the environment for PSA modeling The work involved under this project include the reliability analysis of safety systems and safety support systems, estimation of failure frequency of the initiating events and modeling of accident sequences towards giving the statement of core damage frequency (CDF) for the plant. Event trees analysis had been used to study the response of the plant to various initiating events whereas fault tree had been used in the modeling of safety system failures. The scope of this work extends beyond Level 1 PSA study to include the limited scope Level 2 PSA study to give the likelihood of releases, during the postulated LOCA scenario, to the member of public. The uncertainty analysis was carried out at system level as well as at CDF level to account for possible data and modeling error. The sensitivity analysis was performed to check the affect of major assumptions and critical system parameters on the result of this analysis. The results of this analysis include the statement of CDF and important accident sequences for the plant. The major steps involved in the PSA of Dhruva reactor are as follows:

2.1 Selection of Initiating Events

The main objective of this task was to generate a list of Initiating events (IEs), as complete as possible. The approaches used for selection of IEs are as follows:

- Engineering evaluation,
- Reference to previous list of initiating events considered in the PSA studies,
- Deductive analysis, and Operational experience

Most of the initiating events for this study were identified based on the references to already existing lists of initiating events. This include plant specific documents like safety analysis report for Dhruva reactor and generic source like, Canadian NRU reactor, ORNL High Flux Research Reactor, CANDU Power reactor IEs list. List of initiating events selected for PSA study of Dhruva recator was given below:

- Loss of Off-site Power (LOOP)
- LOCA major—Severance failure of large pipes in Isolatable region
- LOCA major—Severance failure of large pipes in non—Isolatable region
- LOCA major—Severance failure of small pipes in Isolatable region

- LOCA minor—non—Severance failure of large pipes in Isolatable region
- LOCA minor—non—Severance failure of large pipes in non-Isolatable region
- LOCA minor—non—Severance failure of small pipes in Isolatable region
- Loss of Regulation Accident (LORA)
- Severance failure of ECW line in reactor building (non-isolatable portion)
- Non-Severance failure of ECW line in reactor building (non-isolatable portion)
- Reactor trip
- Failure of PW/ECW line in Service Building
- Compressed Air Failure
- Loss of Cooling during Fuel Handling
- Failure of moderator inlet line (Non isolatable portion)

Grouping of initiating event was carried out in such a way that all the events in the same group impose essentially same success criteria on safety systems and safety support systems as well as special condition (challenges to the operator, automatic plant responses etc.) and thus can be modeled using the same event analysis.

2.2 Identification of Safety Functions and Safety System

Safety system and safety function for Dhruva reactor was identified which are important against the core damage. The most usual element of an event sequence model is failure/success of safety system. The system failures are modeled using techniques like Failure Mode Effect Analysis (FMEA), fault tree, Markov modeling and reliability block diagram. System modeling generates the logic model for safety system unavailability. All the five safety system which includes safety support systems for Dhruva reactor is as follow:

- Emergency Cooling System or Shutdown Cooling System (ECS)
- Reactor Protection System (SOR & BSS)
- Emergency Core Cooling System (ECCS)
- Emergency Power Supply System (Class III, Class II and Class I Power supply)
- Containment Isolation and Emergency Exhaust System (CIEE)

Unavailability of safety systems were calculated and are given in Table 1:

2.3 Accident Sequence Modeling

This task comprised of modeling of potential accident scenario with respect to the response of the plant. ‘Event Tree’ analysis methodology was used for accident sequence modeling. In PSA of Dhruva, accident sequence modeling involved construction of event trees with 17 initiating event, safety systems, and human response. The total number of accident sequences identified through Event Tree Analysis was ~ 80. However, using probabilistic criteria and analytical assessment,

Table 1 Safety system/function unavailability

S.N.	Safety system/function	Identification code	Unavailability
1	Primary shutdown system	SOR	9.5×10^{-5}
2	Backup shutdown system	BSS	7.35×10^{-4}
3	Class III power	Class3	3.3×10^{-3}
4	Class II power 150 kVA	ClassII-150kVA	1.5×10^{-4}
5	Class II power 20 kVA	ClassII-20 kVA	6.0×10^{-6}
6	Class I power	Class I	3.0×10^{-5}
7	Non-recovery of Class IV power in 3 h	CLASS4-NR	2.1×10^{-2}
8	Emergency core cooling system (Phase II)	ECCS	4.8×10^{-4}
9	Human error in injection of OHST water	HE-ECCS	3.6×10^{-2}
10	Emergency (or shut-down) cooling system	ECS	2.0×10^{-5}
11	Containment isolation & emergency exhaust	CI&EE	1.1×10^{-2}
12	Human error in isolating ECW line in sub-basement (non-severance failure of ECW line)	HE-NS-ECW	1.0×10^{-3}
13	Human error in isolating ECW line in sub-basement (severance failure of ECW line)	HE-S-ECW	0.37
14	Diesel generator system for OHST	OHST-DG	7.0×10^{-3}

only few accident sequences were identified which are likely to result in varying degree of core damage.

2.4 Accident Sequence Quantification

Quantification of the accident sequences was done by attaching frequency to the initiating event and failure probabilities/unavailability to the safety system along with human error probabilities.

Uncertainty analysis was carried out to measure uncertainty in PSA results. Two approaches were adopted in estimating the uncertainty bounds for various failure rate data. The first approach comprised of estimating the upper and lower bound using the Chi-Square distribution and F-distribution as the case may be or Bayesian approach when the data from two sources are to be combined. The second approach uses Monte-Carlo simulation technique for propagating the uncertainty from component level to top event level in fault tree.

2.5 CDF Statement of Dhruva Reactor

The overall Core Damage Frequency (CDF) value for Dhruva reactor is 4.8×10^{-5}/yr. The median value, lower bound and upper bound of CDF are 3.4×10^{-5}/yr, 2.17×10^{-5}/yr and 7.07×10^{-5}/yr.

Fig. 1 Major contributors to CDF

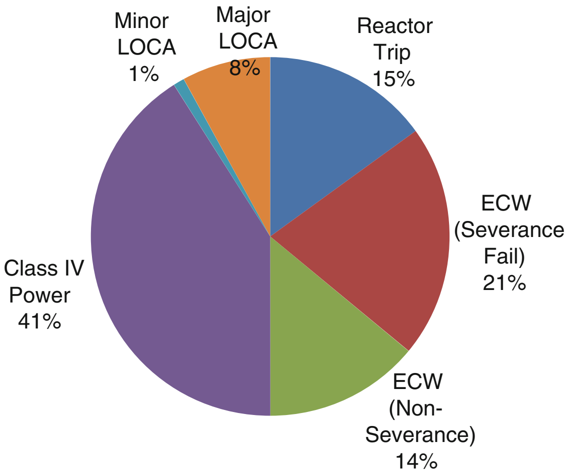
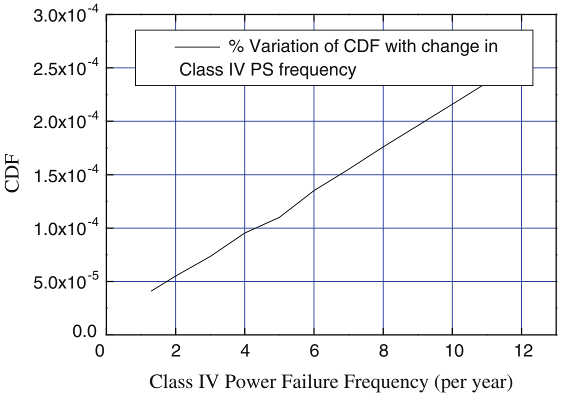


Fig. 2 Sensitivity analysis of Class IV power failure



Around 41 % of the contribution to the net CDF comes from 'Loss of offsite power'. Figure 1 shows the contribution from various postulated accident initiators:

The Initiating Event 'Class IV power failure' has been found to be the greatest contributors to the CDF. The sensitivity analysis performed on this parameter, as shown in Fig. 2, shows that even by considering its frequency as high as 10/yr, the CDF value increases to only 2.25 × 10^{-4}/yr.

3 Level 1 PSA of Cirus Reactor

Cirus was a 40 MWt tank type reactor, having natural uranium as fuel, heavy water as moderator, de-mineralized light water as coolant and graphite as the reflector. Six Boron Carbide shut-off rods was the primary fast shutdown system backed up by the moderator dumping to keep the reactor sub-critical. Seawater was used as the secondary coolant and was the ultimate heat sink. The shutdown cooling system comprised of a once through pass gravity cooling provided by a spherical reservoir called as ball tank. The reactor was housed in a metallic containment building called reactor building. Cirus was commissioned in the year 1960. During the early nineties, the reactor started showing signs and symptoms of ageing which manifested in reduction in availability factor in spite of extensive maintenance efforts. As part of life extension programme for the facility, the ageing studies were performed and the reactor was shut down during the period from 1997 to 2002 to carry out various refurbishment jobs. After the completion of the repair/refurbishing work the reactor systems were commissioned and the reactor achieved criticality on October 2002. Cirus was permanently shut down on 31st December, 2010.

The integrated Level 1 PSA study was initiated in 2005 [4]. Since, this plant had logged in over 45 years of service at the time of commencement of this PSA study, it was decided to use plant specific data to the extent possible for estimating the reliability of various components in general and safety systems components in particular. List of initiating event selected for PSA study is given below:

- Class IV power supply failure
- LOCA major—at outlet side of the reactor core
- LOCA major—at inlet side of the reactor core
- LOCA minor—at outlet side of the reactor core in isolatable region
- LOCA minor—at outlet side of the reactor core in non-isolatable region
- LOCA minor—at inlet side of the reactor core in isolatable region
- LOCA minor—at inlet side of the reactor core in non-isolatable region
- Loss of Regulation Accident (LORA)
- Loss of Flow Accident
- Failure of forward cooling line
- Reactor transients
- Loss of Cooling during Fuel Handling
- Ejection of fuel rod
- Flow blockage
- Failure of Pressure Tube in PWL
- Ball Tank failure

Accident sequence modeling was carried out using Event Tree Technique. The number accident sequences in this analysis were relatively small, i.e. 27 as compared to a typical NPP or even for Dhruva. There are reasons for this as follows: (a) Cirus design was based on single failure criteria hence successful start and operation of the safety system forms part of the design philosophy, (b) the safety system design extensively incorporated passive features which makes the system simple and operation elegant, (c) the coping capacity of the plant in terms of decay heat removal during shutdown mode or accidental mode was very high compared to not only other research reactors but also NPPs, and (d) not many human actions were involved as part of recovery actions due to again relatively large coping time of the plant.

Table 2 CDF statement for cirus reactor

S.N.	Initiating event	Accident frequency (/yr)
1.	Class IV power failure	4.95 × 10^{-6}
2.	Major LOCA	1.01 × 10^{-7}
3.	Minor LOCA	2.52 × 10^{-7}
4.	Reactor transients	1.52 × 10^{-8}
5.	Loss of regulation accident	7.65 × 10^{-9}
	Total CDF	5.32 × 10^{-6}

CDF statement of Cirus comprised of the Median Value, Lower bound and Upper bound of 1.17 × 10^{-6}/yr, 4.12 × 10^{-6}/yr, and 7.71 × 10^{-6}/yr, respectively. Table 2 shows the statement of CDF for Cirus reactor:

4 Level 1 PSA of Upgraded Apsra Recator Project

The Upgraded Apsara Reactor Project is a 2 MWt research reactor. It is a swimming pool type reactor, using LEU as fuel, demineralised water as coolant and moderator and BeO as the reflector material.

Level 1 PSA of Upgraded Apsara Reactor Project has been performed [5]. The analysis has been carried out for full power operation of the reactor considering only the internal initiating events with reactor core as the main source of radioactivity. The available inputs from system Design Basis Reports and plant specific data wherever available from the old Apsara reactor have been used. The initiating events were judiciously screened based on likelihood and consequences and only the selected events are considered for accident sequence modeling. Following initiating events are modeled for detailed quantification:

(a) Loss of Offsite Power Supply
(b) Loss of Coolant Accident
(c) Loss of Regulation Accident
(d) Loss of Shutdown Cooling
(e) Loss of Flow Accident

The system modeling has been carried out using the fault tree approach. The event tree methodology has been employed for accident sequence modeling. The PSA software Fault Tree + version 10.1 supplied by M/s. Isograph has been used for creating model of the plant. The sensitivity analysis has been carried out to check the effect of critical parameters on the results of the analysis. Table 3 shows the statement of CDF for Upgraded Apsara Reactor Project:

The broad insights that were available from the accident sequence analysis are as follows:

- Loss of Offsite Power (CL-IV power) failure is one of the major contributors to the total CDF

Table 3 CDF statement for upgraded Apsara

S. N.	Initiating event	Accident frequency/year	% Contribution
1.	Loss of offsite power	1.08×10^{-6}	68.45
2.	LOCA	7.0×10^{-9}	0.44
3.	LOCA beam tube rupture	4.9×10^{-7}	31.05
4.	Loss of regulation	7.81×10^{-10}	0.049
5.	Loss of S/D cooling	5.79×10^{-14}	Almost nil
	Total CDF	1.58×10^{-6}	

- LOCA contribution to CDF was found very small however, LOCA due to beam tube rupture showed significant contribution.
- Contribution from loss of regulation events is also small.

The results of the analysis demonstrated that the inherent design features of the reactor like passive system employed for shutdown cooling, independent Reactor Protection and Regulation systems make the reactor one of the safe nuclear plant.

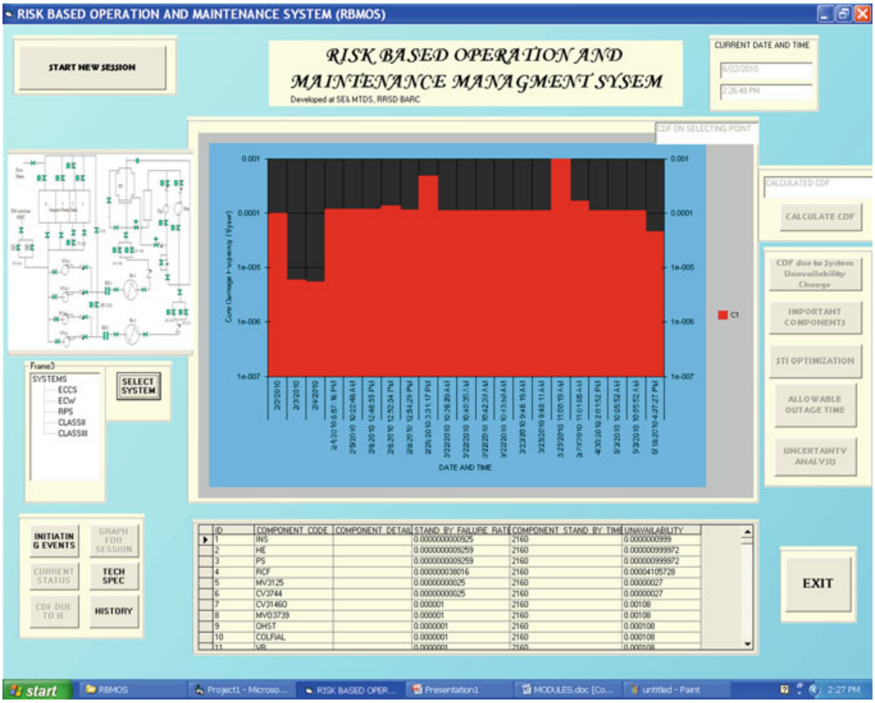
5 Major Applications of PSA

PSA of research reactors provides the safety assessment of the plant. Besides this, PSA have been used for (1) Optimization of Surveillance Test Interval (STI) & Allowable Outage Time (AOT) estimation, (2) Precursor event analysis (incident analysis), and (3) evaluation of Emergency Operating Procedure (EOP).

In BARC, Risk monitor has been developed for the Dhruva reactor. This enables plant operators and schedulers to evaluate the plant risks and problems associated with scheduling and approving outage maintenance activities. Risk Monitor is designed to indicate the risk level for the current plant configuration based on actual configuration, considering risk contributed by each component and system. Risk Monitor helps plant personnel to understand the risks associated with any plant configuration.

Risk Monitor can be used for identification of system & component importance to safety, evaluates the technical specification which includes surveillance test interval, and simulation of accident scenario.

Risk Monitor is a management tool having user friendly Graphical User Interface (GUI) for plant managers and operators with almost all important elements for risk based management such as Core Damage Frequency (CDF) calculation, risk profile graph, system unavailability and IE contribution to CDF, importance analysis, uncertainty analysis, comparison of risk informed surveillance test interval with traditional surveillance test interval, and scope for technical specifications, this system can also facilitates the shutdown maintenance planning and scheduling. Figure 3 shows a screen shot of Risk Monitor:



ID	COMPONENT CODE	COMPONENT DETAIL	STAND BY FAILURE RATE	COMPONENT STAND BY TIME	UNAVAILABILITY
1	INS		0.0000000000925	2160	0.0000000999
2	HE		0.00000000009259	2160	0.000000099972
3	PS		0.00000000009259	2160	0.000000099972
4	RCF		0.00000000038016	2160	0.00004105728
5	MV3125		0.00000000025	2160	0.00000027
6	CV3744		0.00000000025	2160	0.00000027
7	CV31460		0.000001	2160	0.00108
8	MV03739		0.000001	2160	0.00108
9	OHST		0.0000001	2160	0.000108
10	COLFIAL		0.0000001	2160	0.000108
11	VB		0.0000001	2160	0.000108

Fig. 3 Screen shot of risk monitor

6 Conclusion

PSA techniques are applied for Research Reactors in order to reduce their overall risk. The minimal cut sets provided by the PSA can identify the shortest path by which a component failure can propagate, degrade the system and deteriorate the safety. Level 1 + PSA of Dhruva has helped in freezing the final scheme of upgradation of Emergency Core Cooling System. PSA of Upgraded Apsara Reactor project has been carried out in parallel with the basic engineering phase of the project. Therefore, preliminary results have been used to retrofit the design process, thus permitting improvements to the design of system. These improvements have resulted in an effective reduction of the residual risk. PSA has been proved to be a valuable tool to increase the safety level of plant. The main conclusion is that it is possible to effectively reduce the risk of Research Reactors if the basic design process and the PSA proceed in parallel, eventually converging to a reactor where no external emergency plans are necessary.

References

1. U. S. Nuclear Regulatory Commission (1975) Reactor safety study: an assessment of accident risks in U.S. commercial nuclear power plants, WASH-1400 (NUREG-75/014)
2. AERB Safety Guide (2007) Consenting process for nuclear power plants and research reactors. AERB/NPP & RR/SG-G-1
3. Varde PV (2002) Report on Level 1 + Probabilstic Safety Assessment of Dhruva
4. Varde PV (2009) Report on Level 1 Probabilstic Safety Assessment of Cirus
5. Varde PV (2012) Report on Level 1 Probabilstic Safety Assessment of Upgarded Apsara Reactor Project
6. Saraf RL (2006) Evolution of PSA activities in DAE, India, reliability, safety & hazard advances in risk-informed technology. Narosa Publishing House, New Delhi, pp 410–414

Part VI
Reliability Analysis and Modeling

Reliability Model of a Safety System with an Imperfect Tester

Pramod Kumar Sharma and A. John Arul

Abstract High reliability digital systems designed for safety applications employ external or built in test and surveillance systems. The overall reliability is determined by the combined reliability of the safety system and that of the testing system. In this study we derive an expression for the overall system unavailability using Markov model and from that derive an approximate formula that could be used in fault tree analysis of larger systems. The approximate expression is validated by numerical case studies as applied to solid state voting logic with online fine impulse testing system (SLFIT) used in the shutdown system of a nuclear reactor.

Keywords Digital system reliability · Automated testing · Markov modeling

1 Introduction

Frequent testing of the safety systems either manually or by automatic means is one of the many ways available to achieve required reliability. The reliability models often used for calculating the reliability of such systems assume (i) perfect testing, i.e., the system reliability is restored to one immediately after the completion of the test, (ii) The faults that occur in the system is covered 100 % by the test and (iii) The tester or the testing system does not fail. We examine in this report the reliability of the system, when the 3rd condition is not applicable. Modelling the cases corresponding to condition (i) and (ii) are relatively easy and not discussed.

P.K. Sharma (✉)
IGCAR, 107 CDO Building, Kalpakkam 603102, India
e-mail: pramodonline2004@gmail.com

A. John Arul
IGCAR, 111H CDO Building, Kalpakkam 603102, India
e-mail: arul@igcar.gov.in

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_37

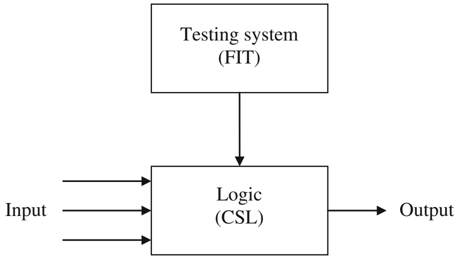
2 System Description and Function

The system considered for study is the voting logic with an automated testing system known as Fine Impulse Tester (FIT). This is typical of a system used in the shutdown system of a nuclear power plant for example Prototype Fast breeder Reactor (PFBR). Each shutdown system consists of a Reactor Protection System (RPS), Actuation System (AS) and safety support systems. RPS consists of instrumentation, i.e., sensors to monitor plant parameters, analogue signal processing circuits, SCRAM logic, SCRAM switches (power gates) and power supply [1]. Actuation System (AS) consists of absorber rods (AR), electromagnets and drive mechanisms to drop or drive the absorber rods into the core.

2.1 Role of Scram Logic

The role of Scram logic is explained here. One type of the SCRAM Logic employs conventional solid state logic with online fine impulse testing (SLFIT). It is built using programmable logic devices (PLD). SLFIT consists of signal conditioners, safety logic core, output stage and annunciation. FIT logic is designed to check unsafe and safe faults of SLFIT apart from the self-diagnostic tests. Scram logic circuit used in SDS1 is essentially of digital signal processing involving voting logic, where normal working condition is coded as high voltage state and trip condition is represented by zero voltage state. System stuck at 1 fault can be revealed only by Fine Impulse Test (FIT). In SLFIT, 2/3 voting for a single parameter is adopted. When two or more inputs are 0, output would be 0. Therefore system will fail only if there are failures in two or more inputs or the voting logic itself fails. The total voting logic system has two parameters. Since any stuck at 1 or 0 failures will not be revealed, they are tested by a FIT circuit as shown in Fig. 1. There is no scram on FIT fault. For the purpose of analysis FIT is treated as one component and Control Safety Logic (CSL) is treated as another component with constant failure rate.

Fig. 1 Model block diagram of SLFIT



3 Reliability Modeling

The system described in Sect. 2 is to be modeled as Markov state space model. The possible system status of the scram/voting logic is shown in Table 1. The first column denotes system states labelled 1, 2, 3 and 4. The component states are 0 or 1. 0 is component failure state and 1 is component success state. The resulting system condition is given in last column. For system state 4, both testing system and scram logic are in failed state. This results in system failure. For system state 3, scram logic is in failed state and testing system is in working state. This also results in system failure. For system state 2, testing system is in failed state and scram logic is in working state, which is considered successful operation of the system. For system state 1 both testing system and scram logic is in working state, which is successful operation of the system. If the scram logic is working then system is safe to operate even though diagnostics is not available. Let, λ_S be the Scram logic failure rate, λ_D the diagnostic system failure rate, and μ_1 and μ_2 are their respective repair rates. Further it is assumed that repair time for scram logic is less compared to diagnostic system's repair time.

The probabilities P_1–P_4 represent the probabilities of finding the system in their respective states 1–4.

3.1 Markov Process

A random process whose future probabilities are determined by its most recent values is called Markov process. A stochastic process x(t) is called a Markov if for every n and $t_1 < t_2 < \ldots .t_n$, we have $P\{x(t_n) \leq x_n | x(t_{n-1}),\ldots,x(t_1)\} = P\{x(t_n) \leq x_n | x(t_{n-1})\}$ [2]. Under the Markovian assumption the system consisting of the SCRAM logic and the testing system are modelled as a collection of states with the transition rates between them as explained in the next section.

Table 1 Logic state

State	Testing system	Scram logic	Result	System safety
4	0	0	0	Unsafe (longer time)
3	1	0	0	Unsafe (shorter time)
2	0	1	1	Safe
1	1	1	1	Safe

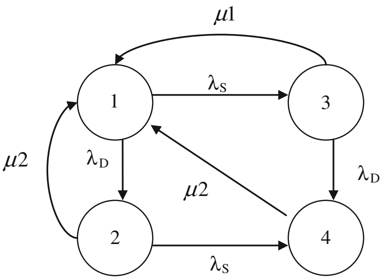


Fig. 2 State transition diagram

3.2 State Transition Diagram

A two-component system has only four possible states, those enumerated in Table 1. The numbers within circle represent system states. The transitions between these states are indicated by directed arcs, labelled with the corresponding transition rates as shown in Fig. 2.

4 Modeling Equations and Steady State Solution

The failure rates λ_D and λ_S are for diagnostics and scram logic respectively. Since λ_S Δt is the probability that SCRAM logic will fail between time t and t + Δt, given that it is operating at t (and similarly for λ_D), we may write the net change in the probability that the system will be in the state 1 as

$$P_1(t+\Delta t) - P_1(t) = -\lambda_D \Delta t.P_1(t) \; - \lambda_S.\Delta t.P_1(t),$$

if the repair rates are neglected.

Or in differential form,

$$\frac{dp_1}{dt} = -\lambda_D.P_1(t) - \lambda_S.P_1(t) \tag{1}$$

Following Eq. (1) the probabilities of finding the system in the respective states considering all transitions (failure and repair) are obtained as the following four equations. The total probability, i.e., sum of P_1–P_4 is 1, can be verified by adding the right hand side of the equations from (2) to (5)

$$\frac{dp_1}{dt} = -(\lambda_S + \lambda_D)p_1 + \mu_1 p_3 + \mu_2 p_2 + \mu_2 p_4 \tag{2}$$

$$\frac{dp_2}{dt} = -(\lambda_S + \mu_2)p_2 + \lambda_D p_1 \tag{3}$$

$$\frac{dp_3}{dt} = -\,(\lambda_D + \mu_1)p_3 + \lambda_S p_1 \tag{4}$$

$$\frac{dp_4}{dt} = -\mu_2 p_4 + \lambda_D p_3 + \lambda_S p_2 \tag{5}$$

Under steady state conditions the above equations could be written as, (setting $\frac{dp}{dt} = 0$)

From Eqs. (3) and (4)

$$p_1 = p_2(\lambda_S + \mu_2)/\lambda_D;\; p_1 = p_3(\lambda_D + \mu_1)/\lambda_S;$$

$$p_2 = p_3(\lambda_D + \mu_1)/\lambda_S.\, \lambda_D/(\lambda_S + \mu_2);$$

From Eq. (5)

$$-\mu_2 p_4 + \lambda_D\, p_3 + \lambda_S\, p_2 = 0$$

And from Eq. (2)

$$\lambda_D p_3 + \lambda_S p_2 - \mu_2 + \mu_2(p_1 + p_2 + p_3) = 0$$

Solving these equations give exact solution as

$$p_1 = \mu_2/\,(\mu_2 + \lambda_S(1 + (\lambda_D + \mu_2)/\,(\lambda_D + \mu_1)));\tag{6}$$

$$p_2 = \mu_2/\,(\mu_2 + \lambda_S(1 + (\lambda_D + \mu_2)/(\lambda_D + \mu_1))).\, \lambda_D/(\lambda_S + \mu_2);\tag{7}$$

$$p_3 = \mu_2/\,(\mu_2 + \lambda_S(1 + (\lambda_D + \mu_2)/(\lambda_D + \mu_1))).\, \lambda_S/(\lambda_D + \mu_1);\tag{8}$$

$$p_4 = 1/\,(\mu_2 + \lambda_S(1 + (\lambda_D + \mu_2)/(\lambda_D + \mu_1))).\, \lambda_D\lambda_S/(\lambda_S + \mu_2);\tag{9}$$

5 Success Criteria

For the system to be in working condition scram logic should be working even though diagnostics is in failed condition, corresponding to state 1 & state 2, as identified in Table 1. The remaining states, state 3 & state 4 are failure states.

The probability of failure on demand (PFD) would be the sum of the probability of finding the system in states 3 and 4.

i.e., PFD = $P_3 + P_4$, and success probability would be $P_1 + P_2$.

6 Approximate Solution

The PFD or equivalently the unavailability of the system is given by $P_3 + P_4$ as shown in Eqs. (8) and (9), which are not convenient for frequent use. Therefore we try to make it simple by making the following assumptions. It has been assumed that repair time for diagnostic is much larger than repair time for scram logic. i.e. $T_1 \ll T_2$ (i.e. $\mu_1 \gg \mu_2$) where T_1 cT_2 is for diagnostic and $\lambda \ll \mu$. This means that failure rate of Scram logic and Diagnostic are much less than their repair rate.

Under these considerations approximate solution would be

$$p_1 = 1/(1+\lambda_S T_1) \cong 1; \tag{10}$$

$$p_2 = \lambda_D/(\lambda_S+\mu_2) = \lambda_D T_2/(1+\lambda_S T_2); \tag{11}$$

$$p_3 = \lambda_S/(\lambda_S+\mu_1) = \lambda_S T_1; \tag{12}$$

$$p_4 = \lambda_S\,\lambda_D\,T_2^2; \tag{13}$$

$$\text{Unavailability} = \mathrm{P}_3 + \mathrm{P}_4 = \lambda_S\,T_1 + \lambda_S\,\lambda_D\,T_2^2 \tag{14}$$

7 Numerical Validation

The Eqs. (2–5) have been solved numerically as a function of time using ISOGRAPH Reliability Software [3] and compared with steady state approximate and exact solution. Exact solution is given by Eqs. (6–9) and steady state approximate solution by Eqs. (10–13). Equation (14) represents unavailability of the system in steady state condition. To enable this comparison typical values of the parameter used are given as $\lambda_S = 3\text{x}10^{-6}/\text{h}; \lambda_D = 2.45\text{x}10^{-7}/\text{h}$. Results are comparable as shown in Table 2. It is evident that for $T_1 \ll T_2$ and $\lambda \ll \mu$ (above said approximation) the calculated value matches exact value.

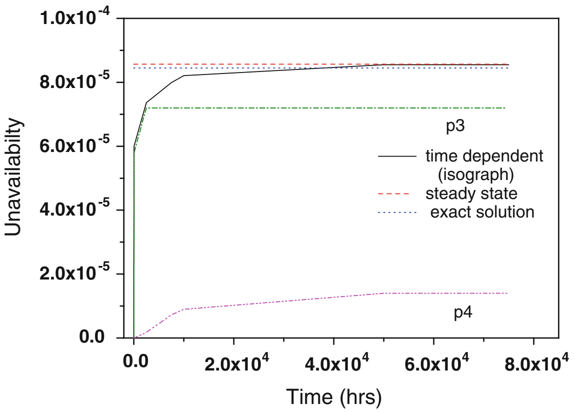
Deviation = (Markov-approximated solution)/Markov*100

All solution with different procedures have been calculated and shown in Fig. 3. Solution with respect to steady state approximation is 8.57E-5 and with exact calculation it is 8.45 E-5 with T_1 = 24 h and T_2 = 4320 h. Also variation of failure

Table 2 Numerical validation

T_1 (h)	T_2 (h)	Markov process		Approximated solution		Deviation (%)
		P_3	P_4	P_3	P_4	(P_3 + P_4)
2	720	6.0E-6	3.8E-7	6.0E-6	3.8E-7	1.57E-02
2	2160	6.0E-6	3.2E-6	6.0E-6	3.4E-6	−2.28E+00
2	4320	5.9E-6	9.2E-6	6.0E-6	1.4E-5	−2.99E+01
8	720	2.4E-5	3.8E-7	2.4E-5	3.8E-7	1.64E-02
8	2160	2.4E-5	3.2E-6	2.4E-5	3.4E-6	−7.34E-01
8	4320	2.4E-5	9.2E-6	2.4E-5	1.4E-5	−1.36E+01
24	720	7.2E-5	3.9E-7	7.2E-5	3.8E-7	1.66E-02
24	2160	7.2E-5	3.3E-6	7.2E-5	3.4E-6	−2.26E-01
24	4320	7.2E-5	9.2E-6	7.2E-5	1.4E-5	−5.62E+00
720	720	2.2E-3	7.6E-7	2.2E-3	3.8E-7	−4.47E-01
720	2160	2.2E-3	4.3E-6	2.2E-3	3.4E-6	−4.22E-01
720	4320	2.2E-3	1.1E-5	2.2E-3	1.4E-5	−5.78E-01

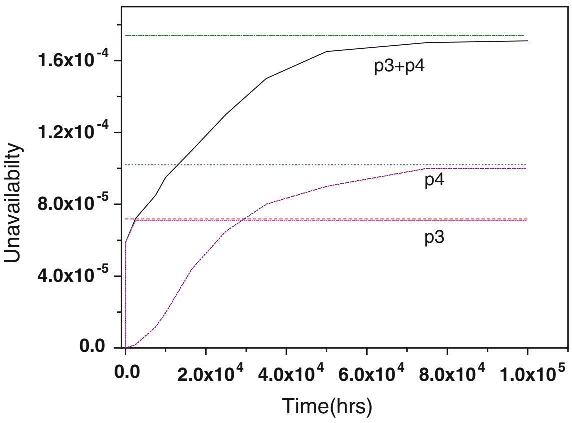
Fig. 3 Unavailability of tested system as function of time (T_1 = 24 h, T_2 = 4320 h)



probability with respect to individual state probability has also been shown in the Fig. 3. The approximate solution is comparable with transient solution within 5 % error.

Figure 4 shows unavailability of the system with same value of failure rate but different values of repair time as T_1 = 24 h and T_2 = 12,000 h. The graph shows the approximate value of unavailability of each system. Fig. 4 indicates that, when the testing interval for the supervising system is $\sim (\frac{rT_1}{\lambda_S})^{\frac{1}{2}}$, where $r = \frac{\lambda_S}{\lambda_D}$ the failure probability contribution from the second term in Eq. (14) is significant.

Fig. 4 Unavailability of tested system as function of time ($T_1 = 24$ h, $T_2 = 12000$ h)



8 Conclusion

An approximate formula for the calculation of unavailability of an imperfectly tested system has been worked out, by taking as an example the SCRAM logic circuit together with testing system typically used in a nuclear reactor. The approximation has been validated by numerical solution for the transient case. The formula derived will be useful in the Fault Tree/Event tree analysis of safety systems.

Acknowledgments The author acknowledges the encouragement and guidance received from Director RDG and Director IGCAR.

References

1. Ramakrishnan M, John Arul A, Bhuvana V, Nagraj CP (2008) Reliability analysis of shutdown system, PFBR/66300/DN/1012
2. Ebeling CE (2000) Reliability and maintainability engineering. Tata McGraw-Hill Edition, New Delhi
3. Isograph Reliability Workbench 10.2. http://www.isograph.com/software/reliability-workbench/

Rail Breaks—An Explorative Case Study

Peter Söderholm and Bjarne Bergquist

Abstract Rail breaks are safety critical failures within railway that may result in derailment, but also delays and cancelled trains. Maintenance is important to both manage the causes of rail breaks and to reduce their unwanted consequences. The purpose of this study is to explore the relationship between maintenance practice, rail breaks and their consequences, to achieve an increased understanding of the rail break phenomenon and promote continuous improvement. To fulfil the purpose, an explorative case study at Trafikverket (Swedish transport administration) was performed. The empirical data was collected from databases that contains information about preventive and corrective maintenance, as well as traffic and traffic disturbances related to rail breaks. The analysis was founded on theories from the three fields of time series analysis, reliability analysis of repairable systems, and multivariate data analysis. The findings of the study support an increased understanding of the process of rail break development and occurrence, but also related maintenance efforts.

Keywords Rail break · Railway · Time series · Reliability · Multivariate · Repairable system · Sweden

1 Introduction

There are extensive work on rail breaks and their maintenance (see, e.g. [1–5]). However, when making operational decisions about the management of rail breaks there are less contributions (see [6], as one exception). For example, at the event of

P. Söderholm (✉)
Trafikverket, Box 809, 971 25 Luleå, Sweden
e-mail: peter.soderholm@trafikverket.se

B. Bergquist
Luleå University of Technology, Universitetsvägen 1,
971 87 Luleå, Sweden
e-mail: bjarne.bergquist@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_38

a postponed track renewal, there are no pragmatic cause-effect relationship between the rail degradation process and associated maintenance actions that can be used to estimate the risk for rail breaks and balance it with increased maintenance efforts and traffic restrictions.

This study is not intended to result in a decision support tool regarding rail breaks and their maintenance, but to give an increased understanding of the rail break phenomenon and suggest an analysis toolbox that can be used to achieve some decision support in a specific situation. However, it is expected that further work can aim at establishing some cause-effect relationships to be used as a supporting tool when making decisions regarding maintenance related to rail breaks.

2 Method and Material

To fulfil the purpose, an explorative case study of the occurrence of potential and actual rail breaks in Sweden is performed. This selection was based on the explorative nature of the study, but also by systematic selection criteria described by [7], i.e. type of research question, no required control over behavioural events, and focus on contemporary events.

Empirical data was collected through interviews and data bases. Quantitative data related to rail breaks was obtained by using Trafikverket data warehouse LUPP (based on SAP Business Objects Web Intelligence). The sources of this data are the inspection and fault reporting systems of Trafikverket (i.e. Bessy and Ofelia respectively). The data warehouse supports business intelligence, and it is possible to connect information about delays to the occurrences of rail breaks. By usage of LUPP it is also possible to extract information about tonnage and train passages for different parts of the rail network during different time periods. For the data used in this study there are three different LUPP-reports created; i.e. one for the result of non-destructive test (NDT) inspections (i.e. the indication of potential rail breaks), one for actual rail breaks (containing the detection of rail breaks by inspections or any other source and related delays), and one for the amount of traffic (i.e. tonnage and train passages).

The reports for potential rail breaks (results from the NDT inspections), actual rail breaks (from inspections and other detections), and tonnage were exported from LUPP to Excel. In Excel, the potential and actual rail breaks were related to each other as well as to the tonnage data through common denominators related to temporal (e.g. year and week) and spatial (e.g. track section) data.

Some analyses were made in Excel (e.g. with the aid of Pivot tables), while some additional analyses were made in the statistical software Minitab. The performed analysis can be seen as consisting of three parts, i.e. time series analysis, repairable system analysis and multivariate data analysis. The theories that acted as a foundation for these three different analysis parts are described in [8–10] respectively. The analysis may also be seen as performed in three steps starting with univariate analysis, via bivariate analysis and ending with multivariate analysis. This is a common approach when performing explorative analyses, see e.g. [10]. However, the focus of this paper is on the bivariate and multivariate steps.

From a reliability point of view, each pair of rails at a track section is considered as a repairable system. A repairable system is a system where component parts are repaired or replaced. When a component failure occurs the entire system is not scrapped. Rather, isolated repairs are made to restore the system to working order. An actual rail break is corrected by replacement of a part of the rail or by welding. A potential rail break is supposed to be corrected before it is developed into an actual rail break, e.g. by grinding or replacement of part of the rail (5 m ≤ length ≤ 10 m). Both these maintenance actions will improve the reliability of that specific part of the rail to some degree. However, these maintenance actions will individually, due to their limited spatial propagations, not significantly affect the overall reliability of the whole track section.

The logic of the analysis of this study is based on the bow-tie model, see e.g. [11]. The unwanted event of a rail break has both causes and consequences. The major cause of a rail break is seen as the amount of traffic, which drives the degradation process. In order to avoid rail breaks, different preventive maintenance efforts are performed, e.g. NDT inspections and rectification of inspection remarks before a rail break occurs. However, when a rail break occurs, the preventive maintenance efforts have failed to act as a barrier. The event of a rail break will in turn result in some unwanted consequences, e.g. train delays. In this case, corrective maintenance may act as a barrier to reduce the consequences of a rail break, i.e. through a good maintenance support performance manifested by short administrative and logistic delay times.

Extracted data from the inspection system Bessy and the fault reporting system Ofelia covers the whole of Sweden from the year 2000 to the 10th of November 2014, and corresponds to 5822 unique inspection remarks (Bessy, 641 remarks) or failure events (Ofelia, 5181 events) that represent rail breaks.

A total of 37,871 potential rail breaks were extracted from Bessy based on performed NDT-inspections during the time period from week 37 in 2001 to week 47 in 2014. The criticality of the actions procreated by these inspections remarks are; Acute, 135; Week, 458; Month, 26171; Year, 4221; Inspection, 6884; and U, 2.

Tonnage and train passage data retrieved from LUPP covers the whole of Sweden from week one of 2009 to week 50 of 2014 and includes passenger, freight and duty trains. In total, this corresponds to 13,036,136,334 tonnages and 35,726,272 train passages.

3 Results

The results of the analysis is presented and discussed in the spatial domain (national and track section), technical domain (track and Switches and Crossings, S&C) and in the temporal domain (season, month and year, but also tonnage).

3.1 Preventive Maintenance Practice

The preventive maintenance practice of rail breaks mainly follows the logic of a degradation process, where a failure event will result in a faulty state if no maintenance actions is taken to prevent further degradation. In addition to the degradation process, sudden rail breaks can occur due to abnormal events, such as the impact of a flat wheel. These two causes of rail breaks can also interact, where a degraded rail is more vulnerable to the forces applied from the rolling stock. Furthermore, other factors such as climate and installation, as well as production and installation will affect the probability of rail breaks.

Non-Destructive-Testing (NDT) and other maintenance inspections are performed with specified intervals to detect rail degradation and on-going failure events, i.e. defects that can be seen as potential rail breaks. The intervals are mainly based on the traffics' speed and axle loads at that specific part of the infrastructure. Other affecting factors, such as those mentioned above should also be considered. Rail break failures are to be rectified within a stipulated time in order to prevent the occurrence of actual rail breaks [5, 12].

If an actual rail breaks do occur, it can be detected during the preventive maintenance efforts, but also through failure indications in the signalling system, through observations (e.g. by maintenance personnel or train drivers), or in worst case through a derailment.

It is normally assumed that NDT-remarks are rectified within the stipulated time to prevent actual rail breaks. However, an analysis of the NDT-remarks shows that:

- There are NDT-remarks that not are rectified within the stipulated time, but that does not result in any rail break.
- There are NDT-remarks that are not rectified within the time frame of this study.

Regarding the first category of NDT-remarks, a deeper analysis has to be performed to see if these remarks represent an increased rail break risk, or if they indicate an opportunity to improve the maintenance practice, e.g. by reducing their criticality levels or extending the inspection intervals.

The other category of NDT-remarks represents a time truncation of the analyzed data due to time restrictions of the performed study, and should be managed by an appropriate analytic approach. It should be noted that the analyzed data is truncated both in time and failure. Regarding the time truncation, the total age of a particular rail section can be determined to some degree regarding calendar time, and to less degree in metric tonnages. However, the history of the occurrence of potential and actual rail breaks of a specific track section before the data was covered in the inspection system (Bessy) is more difficult to reproduce. At the same time, the rail remains in operation after the time period covered by the analyzed data. In addition, the occurrence of a potential rail break that is corrected before an actual rail break occurs can be seen as time truncated from a rail break perspective. The data is also failure truncated regarding the actual rail break events.

3.2 Some Traffic and Maintenance Variables Related to Rail Breaks

An initial analysis of related variables was performed to identify variables that may affect rail breaks and their maintenance. The Swedish railway system has been divided into five different asset classes based on their criticality. Asset class 1 is the most critical and represents major city areas, while asset class 5 is the least critical and represents lines with little traffic. In addition, marshalling yards are represented by a sixth asset class (98), due to their complexity and uniqueness. Some differences besides the amount of traffic, e.g. due to the type of train are found concerning the traffic for each asset class considering both the tonnage and the number of trains. For example, asset class 1 has more passenger trains with lower weight than asset class 2. Since tonnage is believed to affect the rail degradation more than the number of trains, the tonnage is used as a time measure for degradation in this study. However, when studying the consequences of rail breaks regarding delays and their costs, it might be more appropriate to consider type and number of trains than tonnage. This is due to that a passenger train may generate more societal costs than a freight train. However, these considerations are excluded here.

Besides the traffic density (i.e. a measure of the age of the rail), it is possible that other variables affect the occurrence of rail breaks, e.g. type of rail, sub ground, climate, and temperature. Based on the extracted data it is possible to study the effect of some of these variables. However, it should also be noted that some of these variables also affect the preventive maintenance practice, and thereby affect the data indirectly.

Multivariate data analysis is suitable for studying if some variables are more related to each other by providing similar information about potential and actual rail breaks. Two potentially useful explorative multivariate data analysis techniques are Cluster Analysis (CA) and Principal Component Analysis (PCA). In these techniques all variables are analysed simultaneously without dividing them into response (dependent or Y) or explanatory (independent or X) variables (e.g. as in multiple regression analysis). One result of the performed CA is illustrated in the dendrogram illustrated in Fig. 1. In the dendrogram, four distinct clusters of variables are visible (below the red line).

One cluster of variables (leftmost in Fig. 1) seems to be related to corrective maintenance of rail breaks in track, and contains the three variables; rail break in track, disturbed trains due to rail break in track, and the resulting delays due to rail breaks in track. The second cluster (from the left in Fig. 1) seems to be related to corrective maintenance of S&C and contains the same variables as for track, but also the number of train missions. In both these clusters it is seen that the number of delayed trains and amount of delays are closely related to each other, i.e. for both track and S&C. Hence, either of these variables can probably be used since they seem to contain similar information.

The third cluster (from left hand side in Fig. 1) is related to inspection remarks and rail breaks detected through preventive maintenance inspections, and contains

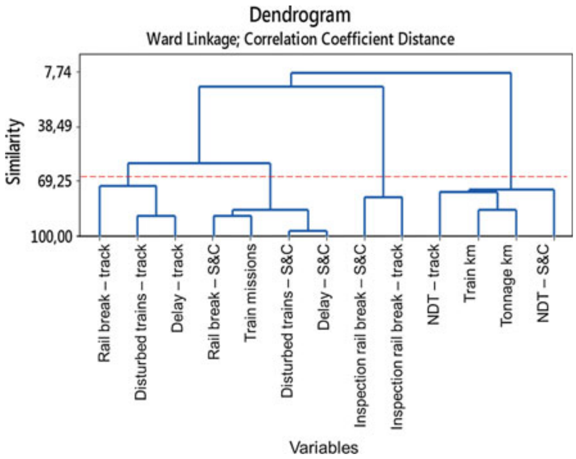


Fig. 1 Dendrogram of the included variables related to potential and actual rail breaks

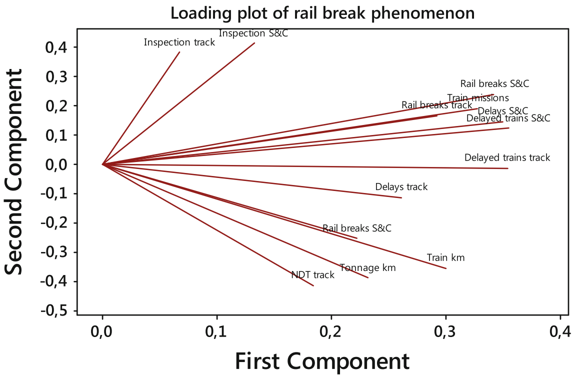
both track and S&C. The forth cluster (rightmost in Fig. 1) contains variables that are related to preventive maintenance performed as NDT-inspection, and also contains the variables train kilometres and tonnage kilometres. The two latter variables are also closely related and seem to contain similar information.

If one wants to reduce the number of clusters from four to three, the two first clusters related to corrective maintenance are closest to each other and can be merged. Hence, if three clusters are used, one cluster can be considered to describe corrective maintenance, another cluster to describe maintenance inspections, and the third cluster to describe preventive maintenance through NDT-inspections. If one wants to reduce the clusters to only two, the corrective maintenance cluster and the preventive maintenance inspections cluster are closest to each other and could be merged, while the NDT-inspections remains as a second cluster. The reason for this is probably that the data that are selected from the inspections and faults represent actual rail breaks, while the NDT-remarks represent potential rail breaks.

As for the dendrogram of variables from the CA, the loading plot of the rail break phenomenon from the initial PCA reveals that the number of variables can be reduced (see Fig. 2). For example, the amount of traffic can either be expressed as tonnage kilometres or train kilometres, as also indicted by the fourth cluster in the CA. The rationale for including kilometres in the measure of traffic is that the length of the track section also is included. The length of the track will affect the probability of a rail break on track section level since the track is a linear asset. However, since S&C are point assets, we consider tonnage in combination with the number of S&C as a proper measure to use to estimate the probability of rail breaks in S&C on track section level. Since the tonnage are believed to affect the degradation of rail more that the number of trains, tonnage kilometres are selected.

As another example, the loading plot reveals that the amount of delays are closely related to the number of delayed trains, where both variables in turn are related to the total number of train missions. This relationship is also illustrated in

Fig. 2 Loading plot of the included variables related to potential and actual rail breaks

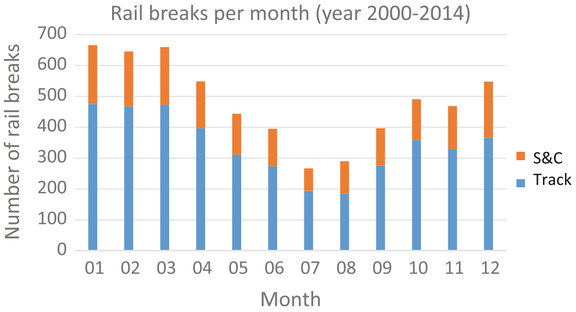


the clusters related to corrective maintenance achieved through the performed CA (see Fig. 1). Since punctuality and delays are two of the most important quality parameter besides safety for the end customers, the amount of delay minutes is selected as a measure of the consequences of rail breaks. Hence, when deciding upon maintenance regarding rail breaks the traffic should be described by both the tonnage kilometres (related to the causes of rail break) and the number of train missions (related to the consequences of rail breaks). The traffic variable of tonnage kilometres will be related to the cause of rail breaks through rail degradation. The traffic variable of train missions will in turn be related to the consequences of rail breaks expressed in delay minutes.

3.3 *Seasonal Influence on Rail Breaks*

If rail breaks demonstrate a seasonality component, knowledge of the seasonality magnitude may influence maintenance planning. The number of rail breaks per month were therefore studied, and there is a seasonal dependence, i.e. rail breaks are more frequent during the winter period than during the summer period (see Fig. 3).

Fig. 3 The number of rail breaks in S&C and track per month for the time period 2000–2014



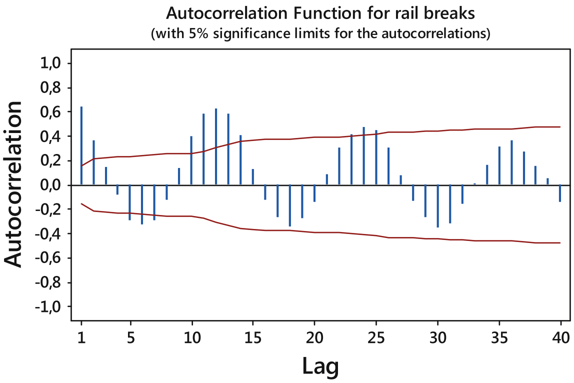


Fig. 4 The ACF for actual rail breaks registered in the fault reporting system

The seasonal pattern of rail breaks from the corrective maintenance data base is also visible as a slowly decaying oscillation of the autocorrelation function, with a seasonality corresponding to 12 months (Fig. 4).

The autocorrelation function (ACF) of the rail break data time series supports the hypothesis that there is a significant seasonal pattern with a yearly cycle. The autocorrelation function reveals that the same month every year (a lag of 12 month) has a positive autocorrelation, i.e. a month with a high (low) number of rail breaks tend to have this every year. In contrast, the autocorrelation at a lag of 6 month is negative, indicating that a month with a high number of rail breaks will be followed by a month with a low number of rail breaks half a year later, and vice versa.

To fit an appropriate time series model to the rail break data and to further test the assumption of a seasonal effect, it may be good to initially compare an additive model with a multiplicative model.

In the additive model, the effects of individual factors are differentiated and added together to model the data. An additive model should be used when the magnitude of the data does not affect its seasonal pattern. In contrast, the multiplicative model should be used when the size of the seasonal pattern depends on the level of the data. This model assumes that as the data increase, so does the seasonal pattern. Most time series plots exhibit such a pattern. In a multiplicative model, the trend and seasonal components are multiplied and then added to the error component. See Figs. 5 and 6.

Considering rail breaks on a national level, it is indicated that a multiplicative model (Fig. 6) is slightly better than an additive model (Fig. 5). The reason for this statement is that the mean absolute percentage error (MAPE), mean absolute deviation (MAD) and the mean squared deviation (MSD) statistics are larger in the latter case, which also is supported by a residual analysis. The residuals of the multiplicative model are more independent and normally distributed with a constant variance than for the additive model. Hence, there is both a linear trend and a seasonal pattern present in the rail break data. The linear trend indicates that the

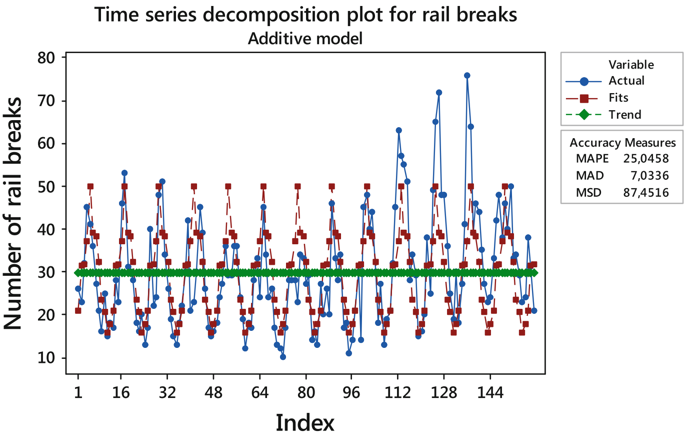


Fig. 5 Additive time series model of rail breaks registered in the fault reporting system

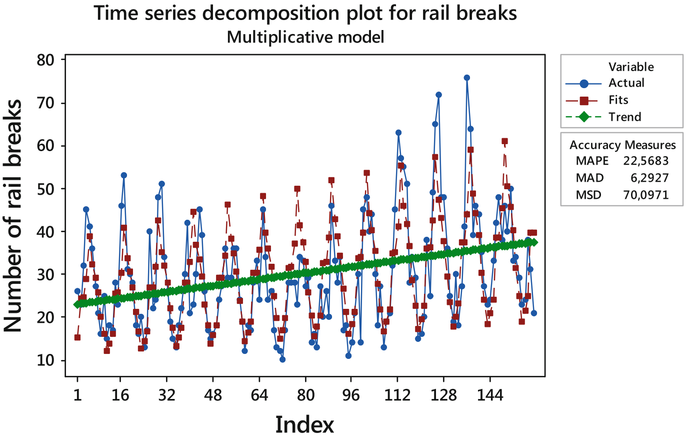
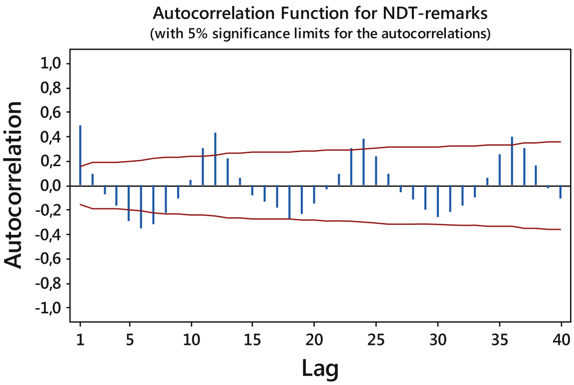


Fig. 6 Multiplicative time series model of rail breaks registered in the fault reporting system

number of rail breaks tend to increases during the time period included in the study. It may be realistic to assume that this increasing trend of rail breaks corresponds to a degrading condition of the rail, i.e. its reliability is decreasing. The seasonal part of the model indicates that there are more rail breaks during the winter period than during the summer period.

The NDT-remarks also display a significant autocorrelation, and a seasonal pattern (see Fig. 7), similar to rail breaks.

Fig. 7 The ACF for potential rail breaks (NDT-remarks) registered in the inspection system on national level



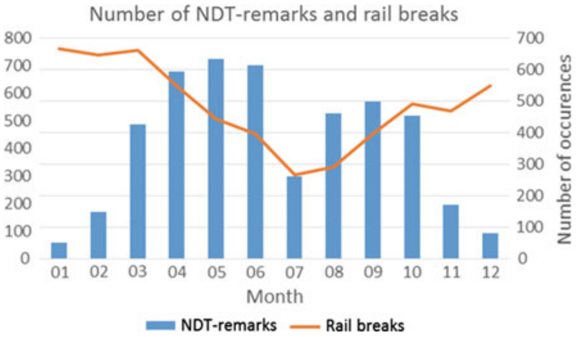
3.4 Relationship Between Potential and Actual Rail Breaks

The relationship between potential and actual rail breaks is interesting from a rail break prevention perspective. The months when NDT inspections are performed and results in inspection remarks are illustrated in Fig. 8. A comparison of the number of NDT-remarks and rail breaks on a monthly basis and a scatterplot may reveal cross correlation. The hypothesis being that NDT-remarks (potential rail breaks) are corrected and thereby prevents the occurrence of actual rail breaks. This comparison is illustrated in Figs. 8 and 9.

The comparison between rail breaks and NDT-remarks indicate that there is a negative correlation at the same month, i.e. a high level of one variable corresponds to a low level of the other. The seasonal influence on the number of rail breaks is also visible, i.e. there tend to be more rail breaks during the winter period than during the summer period. See Figs. 8 and 9.

The cross correlation between the two time series of NDT remarks and rail breaks detected by other means than other types of inspection, is applied to determine if the NDT-series of data leads the rail break series and by how many time periods, or lags. The correlation between the NDT series and the rail break series plus or minus the number of lags (K) is illustrated in Fig. 10.

Fig. 8 Number of actual and potential (NDT-remarks) rail breaks per month on national level



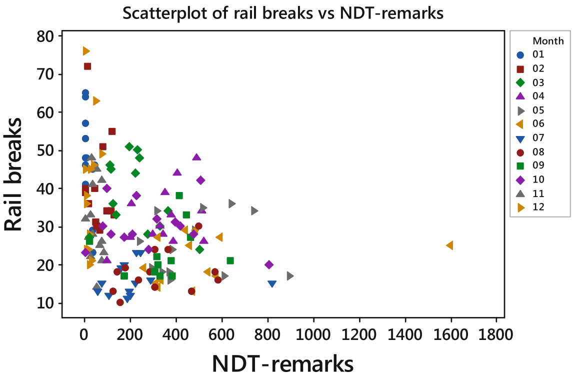


Fig. 9 Scatterplot of actual versus potential (NDT-remarks) rail breaks per month on national level

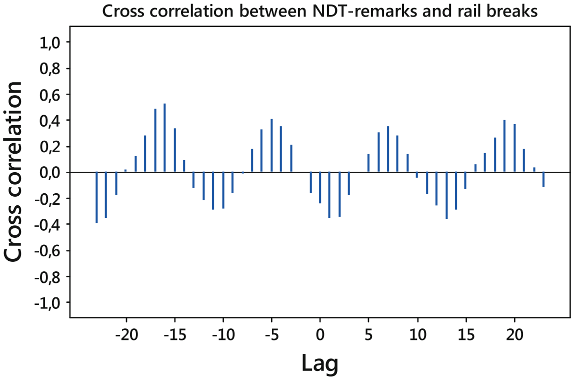


Fig. 10 The CCF for potential (NDT-remarks) and actual rail breaks on national level

The two time series are cross correlated (see Fig. 10). Hence, the outcome of NDT-inspections correlate with the number of rail breaks. The negative cross correlation is present within a lag of 6 months, which indicate that a high number of NDT-remarks will correspond to a low number of rail breaks during the following 6 month. This pattern is also visible in the scatter plot of rail breaks versus NDT-remarks and the monthly stratification of rail breaks and NDT-remarks, where a low level of NDT-remarks corresponds to a high level of rail breaks on a monthly basis (see Figs. 8 and 9).

The cross correlation in Fig. 10 also reflects the seasonal pattern in rail breaks, where rail breaks are more common during the winter period than during the summer period. In addition, this seasonal pattern is probably emphasised by the practice of NDT-measurements, which mainly is performed during the summer period. The NDT-inspections use water as a medium and thereby only can be applied for temperatures down to about −5 °C by adding anti-freeze. Furthermore, winter conditions also contributes as a driver for the development of rail breaks. The latter is due to that the steel of the rail turns brittle at low temperatures, and thereby more sensitive to impacts. The brittleness also leads to notch brittleness, that is, fatigue cracks may propagate much faster when the crack tip is not blunted by plastic deformation rapidly leading to rail breaks. Simultaneously, the steel heat expansion leads to compressive rail stresses during summer, which may close the cracks that are open and more easily spotted during winter when subjected to tensile stresses.

There is a positive cross correlation between the number of NDT-remarks and the number of rail breaks at a lag between 7 and 12 months (see Fig. 10). This means that a high level of NDT-remarks will corresponds to a high level of rail breaks in the time period 7–12 months later. This positive cross correlation may reflect the deterioration process of the rail, which also is indicated by the linear trend of the time series of the rail breaks (see Fig. 6).

However, to get a deeper understanding of the cross correlation, it is necessary to use more advanced time series analyses, and also consider the criticality of the inspection remark, but possible also to stratify on a track section level.

3.5 Rail Breaks in Different Asset Types

Another stratification is to consider the two different types of assets where rail breaks occur, i.e. track and S&C. One reason for this stratification is that there are different regulations and administrative responsibilities connected to the two different asset types. This stratification is illustrated by the performed CA and PCA, where clusters of variables were identified (see Figs. 1 and 2). The rationale for this stratification is also that track is a linear asset, while S&C are point assets. From a maintenance perspective, the differentiation between point assets and linear assets is depending on the criticality that the length of the asset has. The length of point assets, such as S&C, is not critical for their maintenance. When dealing with a point asset, maintenance actions are not assigned to a particular length of the asset, but rather to the entire asset or to some of its indenture levels (included items). However, a linear asset is an asset whose length plays a central role in its maintenance, e.g. railway track. When performing maintenance actions related to linear assets (e.g. NDT-measurement, grinding and welding), it is necessary to be able to define the location of a point or a section along the asset.

In addition to the NDT-measurements, there are other safety inspections, where both potential and actual rail breaks can be detected. Due to the difference between point and linear assets, these safety inspection are probable more effective in detecting rail breaks for S&C than for track.

3.5.1 Track

For rail breaks found in the failure reporting system the ACF also show a seasonal pattern with a 12 month cycle (see Fig. 11).

Hence, with about 6 months lag there is significant negative autocorrelation, while there is a significant positive autocorrelation every 12th month. This means

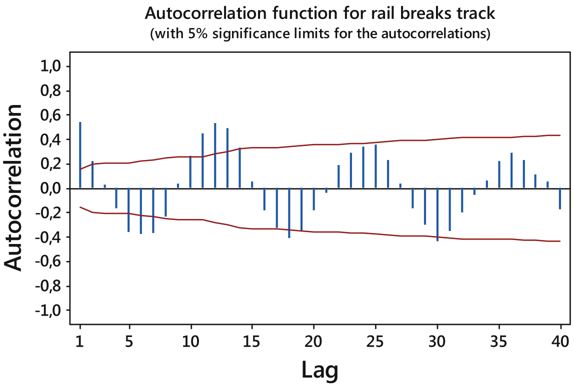


Fig. 11 The ACF for actual rail breaks in track registered in the fault reporting system

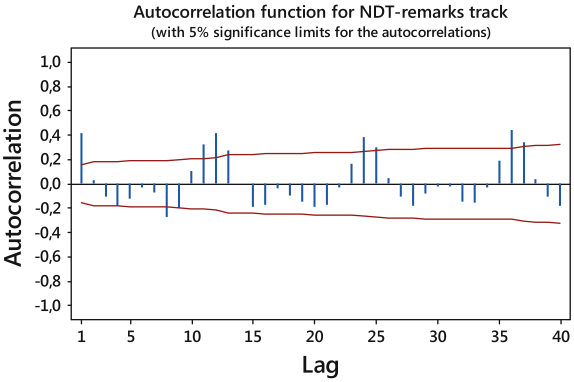


Fig. 12 The ACF for potential rail breaks (NDT-remarks) for track registered in the inspection system

that every specific month each year tend to have the same level of rail breaks every year, while the level of rail breaks is opposite for months with a distance of about 6 months in-between. Hence, the summer months tend to have a lower level of rail breaks while the winter months tend to have a higher level of rail breaks. The same pattern as for rail breaks can be seen in the autocorrelation for NDT-remarks in track, but with a 6 month shift (see Fig. 12).

The CCF between NDT-remarks and rail breaks for track indicates a seasonal pattern with a 12 month cycle, where the correlation is positive during 6 months and negative during 6 months. See Fig. 13.

3.5.2 S&C

The ACF for rail breaks detected in S&C and registered in the fault reporting system (Ofelia) has another pattern compared to the other ACF for the other analysed data. As for the other data, there is a seasonal pattern present, indicating a periodicity of 12 month, which also is reducing with time lag. However, there is no significant negative autocorrelation present. See Fig. 14.

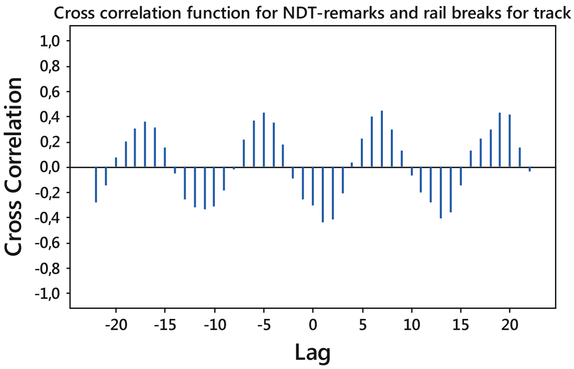


Fig. 13 The CCF for potential (NDT-remarks) and actual rail breaks in track

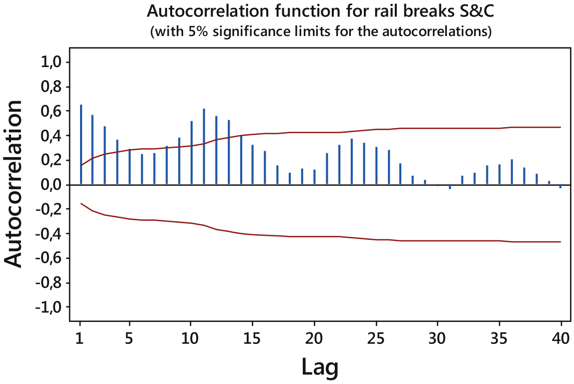


Fig. 14 The ACF for actual rail breaks in S&C registered in the fault reporting system

The reason for the pattern in the ACF for rail breaks in S&C indicates that the time series is not stationary, i.e. neither the mean nor the variance are constant with time. The time series plot of rail breaks for S&C supports this assumption by indicating that both the mean and the variance are increasing with time. In fact, the increasing trend for all rail breaks (Fig. 6) can mainly be attributed to S&C since the time series plot of rail breaks in track indicates a stationary behaviour. Hence, this would indicate that the population of S&C is deteriorating over time with regard to rail breaks.

The ACF for NDT-remarks in S&C indicates a seasonal pattern with significant autocorrelation present. See Fig. 15.

The CCF between NDT-remarks and rail breaks for S&C shows low cross correlation (<0.2), but indicate a seasonal pattern. See Fig. 16.

The CCF between NDT-inspections and rail breaks is about twice as large for track (0.4) as for S&C (0.2), cf. Figs. 13 and 16 respectively. This indicates that the NDT-inspections have more influence on the rail breaks for track than for S&C. This is probably due to the difference in maintenance practice between linear assets and point assets.

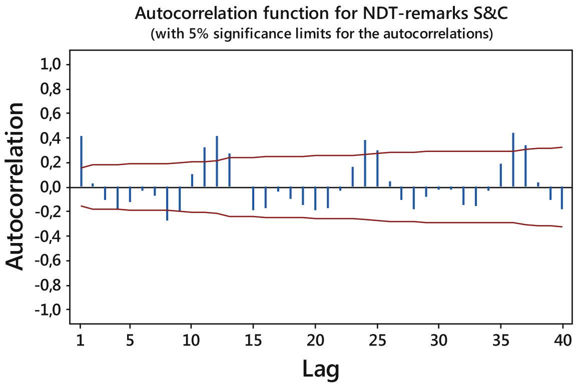


Fig. 15 The ACF for potential rail breaks (NDT-remarks) for S&C registered in the inspection system

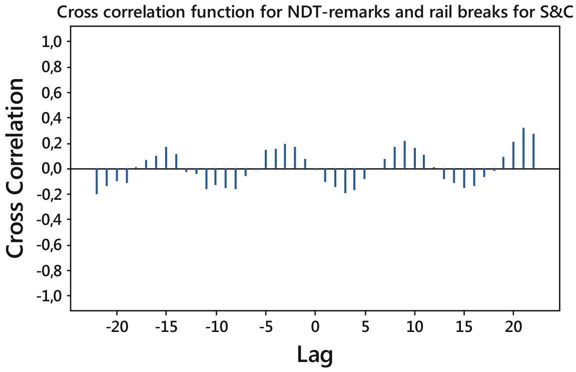


Fig. 16 The CCF for potential (NDT-remarks) and actual rail breaks in S&C

3.6 Rail Breaks at Track Section Level

It may also be interesting to investigate the rail break data on track section level, rather than at the national system level. The reason for this is that a track section often is a part of the rail network that is rather homogenous regarding the amount of traffic that passes through it, at the same time as major maintenance decisions and actions, e.g. track renewal, often is performed at track section level.

A PCA was performed to identify differences between track sections regarding potential and actual rail breaks. This PCA includes a reduced number of variables, as discussed in relation to the performed CA. The PCA includes data about rail breaks from the inspection and fault reporting systems, as well as consequences of rail breaks and amount of traffic related to the track sections. The loading and score plots are visualised in Figs. 17 and 18 respectively.

The first principal component (PC1) seems to mainly be related to traffic characteristics, either as a cause to potential and actual rail breaks, but even more to the consequences of actual rail breaks. Hence, corrective maintenance related to actual rail breaks resulting in delays is also reflected in PC1. See Fig. 17.

Fig. 17 Loading plot of reduced number of variables related to potential and actual rail breaks

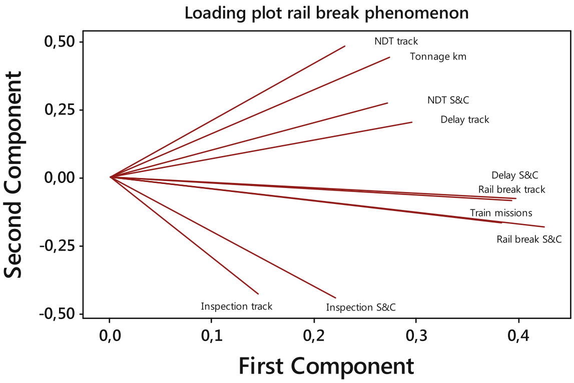
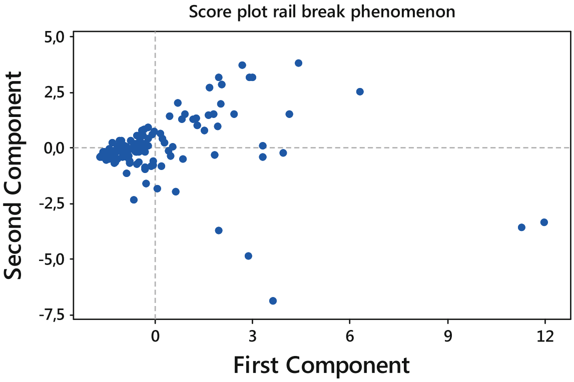


Fig. 18 Score plot of reduced number of variables related to potential and actual rail breaks



The second principal component (PC2) seems to mainly reflect type of preventive maintenance action, where there are two distinct groups with opposite impact (see Fig. 17):

- potential rail breaks detected through NDT-inspections, affected by the amount of traffic that has a positive loading.
- actual rail breaks detected through other types of maintenance inspection neither resulting in any major traffic disturbances nor affected by the amount of traffic, that has a negative loading.

Regarding PC2, it should be noted that the interval of both these preventive maintenance actions are related to traffic characteristics, i.e. at least the tonnage and the speed of trains. Furthermore, machine inspections are probably more cost efficient for linear asset inspection than for point asset inspection compared to manual inspections, due to the differences in detectability of failures.

Track sections that are related to, or affected by, different variables can be identified through a comparison of the loading plot (Fig. 17) and the score plot (Fig. 18).

Two track sections contribute largely to PC1 (i.e. track sections 401 and 601). Furthermore, three track sections contribute negatively to PC2 (i.e. track sections 417, 603 and 902). These track sections also correlate with maintenance inspections of both track and S&C.

In addition, two track sections that have large positive loadings on PC2 and relation to NDT-inspections (i.e. track sections 124 and 912).

The three identified groups of track sections are:

- A group related to NDT-inspections represented by track sections 111, 124, 512 and 912. All these sections have mixed traffic, but the first two sections are single track while the two other are double track.
- A group related to inspections, which consists of track sections 417 (Hallsberg marshalling yard), 603 (from Gothenburg Kville to Gothenburg Skandia harbour) and 902 (Malmö freight station). Hence, these track sections are related to marshalling yards and freight stations, where little NDT-inspections are performed and the geographical area is limited.
- A group related to corrective maintenance actions represented by track sections 401 and 601. These two track sections are in the vicinity of Sweden's two largest cities (Stockholm and Gothenburg respectively) and experience a lot of traffic.

A deeper qualitative analysis of information about the track sections above identified three specific sections for further analysis.

The first track section is number 124 between Boden and Bastuträsk, which for many years has experienced postponed track renewals. At this section, derailments causing extensive traffic disturbances occurred in 2008 and 2013. The track section also has a specific type of rail (Domnarvet 1976–1982) with manufacturing deficiencies, which experiences unwanted vertical crack propagation. The NDT-inspections have been intensified after the derailments and all remarks are rectified.

The second track section is number 417, which is the Hallsberg marshalling yard. The yard has experienced disproportionately many rail breaks and large part of the yard is not inspected through NDT.

Track section 401 is located within the central parts of Stockholm (from Älvsjö via Stockholm C to Ulriksdal), and is thus exposed to a large traffic load. Simultaneously, it has a high number of rail breaks and NDT-remarks.

3.7 Consequences of Rail Breaks

Examples of potential consequences of rail breaks are derailments, delays and cancelled trains, as well as corrective maintenance actions.

3.7.1 Traffic Consequences

Measures of the consequences of a rail breaks include the number of disturbed trains and delay minutes. From January of 2010 until the 10th of November 2014 there are 2318 reported rail breaks (inspection remarks and fault reports), which resulted in 10,722 delayed trains on a right time level of three minutes (RT +3) and 250,819 min of delay (almost 6 months).

The number of disturbed trains correlate positively with the amount of delays (Figs. 19 and 20). A reasonable explanation is that a high number of disturbed trains also would result in a high volume of delay minutes (see also Figs. 1 and 2).

Rail breaks in plain track result in more disturbed trains and delay minutes than rail breaks in S&C (see Fig. 19). This might also be reasonable due to the different characteristics of the two asset types, i.e. plain track is a linear asset, while S&C is a point asset. This will in turn affect the maintenance of respective asset, which also will affect the maintenance support performance of the maintenance organization, e.g. logistic delay times, and thereby traffic disturbances. Hence, depending on how the rail break is detected, it should mainly be possible to initially localize the rail break to a specific S&C. However, for plain track the initial localization of a rail break is probably possible to achieve down to a section block (e.g. through the signalling system). Hence, to perform corrective maintenance of a rail break, the S&C provides a limited area (point) to cover for fault localization, while the track would provide a larger area (distance) to cover in order to localize the rail break.

The positive correlation between the number of disturbed trains and delay minutes differs depending on asset classes (see Fig. 20). For example, the positive correlation decreases with asset class. This indicates that the more traffic (i.e. the higher the asset class), the more vulnerable the traffic becomes to rail breaks (see also Figs. 2 and 3). In addition, the two asset classes 1 and 2 tend to have more disturbed trains than the other asset classes, but also to have a higher volume of delays. This seems reasonable since the amount of traffic is highest at these two asset classes.

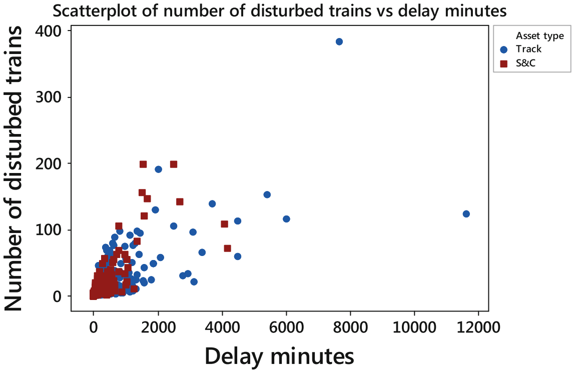


Fig. 19 Scatterplot of number of disturbed trains versus delay minutes for rail breaks in S&C and track

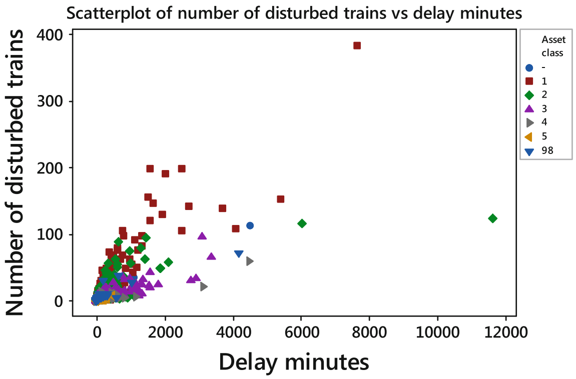


Fig. 20 Scatterplot of number of disturbed trains versus delay minutes for rail breaks in different asset classes

The most extreme values for traffic disturbances can be appointed to rail breaks in tracks of asset classes 1 and 2. Hence, this indicates that a combination of high traffic and a linear asset type has a negative effect on the infrastructure's robustness when considering rail breaks.

3.7.2 Corrective Maintenance

The practices of corrective maintenance tasks at a rail break in track and in S&C are somewhat different (see Figs. 21 and 22). The most common track maintenance task is repair, followed by replacement of unit (Fig. 21). For S&C the order of these tasks are the opposite (Fig. 22). One reason for this difference might simply be that it is easier to change a rail part of the S&C than a part of the rail in plain track. However, when looking at the amount of delay per corrective maintenance action, it

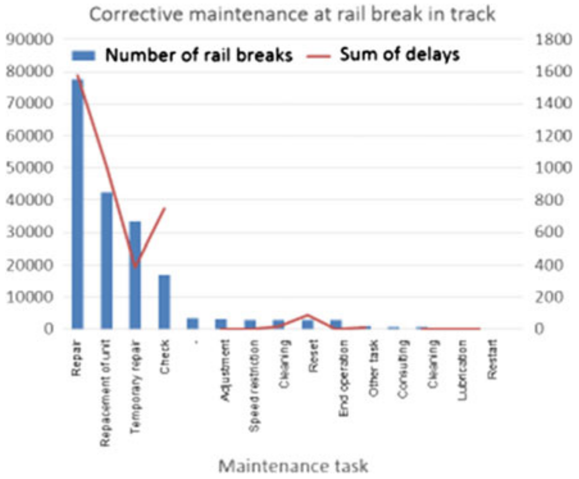


Fig. 21 Corrective maintenance actions of rail breaks in track and related delay

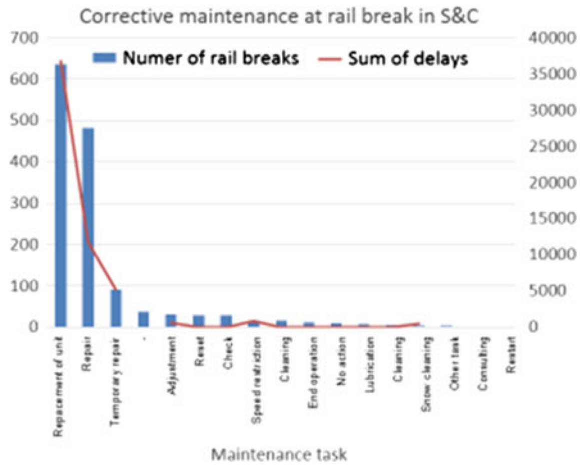


Fig. 22 Corrective maintenance actions of rail breaks in S&C and related delay

is seen that replacement of unit results in more delays than repair, which might indicate that the former action takes more time (see Figs. 21 and 22). This is especially true for S&C, where the replacement of unit corresponds to more delay time per rail break than repair (see Fig. 22).

The third most common tasks for both asset types are temporary repair (see Figs. 21 and 22). This also seems logical, since the purpose would be to minimize the time for corrective action and thereby the traffic disturbances. Once the temporary repair is finished, it is possible to plan for a permanent maintenance action at more convenient time. This practice seems more effective for track than for S&C regarding the amount of delay per maintenance occasion (cf. Figs. 21 and 22).

With one exception for track, the other types of maintenance tasks are few compared to this top three. The exception for track is the fourth most common task, i.e. to check the condition at a rail break (see Fig. 21). This task might be related to the practice to check the rail where a train that is suspected to be damaging the rail has passed a wheel impact detector. The impact detector is for instance triggered by a flat wheel. This practice takes time since the maintainer has to walk along the track and look for damages, which also is indicated by the amount of delay per maintenance action.

4 Discussion and Conculsions

This paper explores the phenomenon of rail breaks on an aggregated level. Some hypotheses have been strengthened and some new ones have emerged. Interesting further research would be to conduct supplementary descriptive and explanatory studies on a more detailed level, to establish more tangible relationships between cause and effect. Based on the performed explorative analysis and its results, the following tentative conclusions can be drawn:

- There is a cross correlation between the number of NDT remarks and the number of rail breaks, which indicates that it is possible to affect the number of rail breaks through altering the practice of NDT inspections. The detection of potential rail breaks (NDT-remarks) and their correction seems to have a positive effect during a time period of about 6 months, where the cross correlation is negative, i.e. a high number of NDT-remarks corresponds to a low level of rail breaks during 6 months.
- There is a seasonal component in the time series data for rail breaks, indicating that there are more rail breaks during the winter period than during the summer period. This seasonal effect is probably strengthened by the practice of performing the NDT inspections, which mainly is performed during the summer period. Nevertheless, to reduce the number of rail breaks, one hypothesis generated by the results is that an increased number of NDT inspections during the winter period, using technology to overcome temperature difficulties may be effective.
- There is a linear trend in the time series data for rail breaks, indicating that the seasonal component increases with time. This linear trend probably reflects that the rail deteriorates within the studied time period. Hence, to reduce the number of rail breaks, an increased number of NDT inspections in response to an increased age of the rail may be effective. Other preventive maintenance initiatives may be to replace the rail earlier and thereby avoid that it becomes more sensitive to rail breaks due to aging effects of the rail material. The aging effect is also indicated in the cross correlation between NDT-remarks and rail breaks, where there is a positive correlation at a lag between 7 and 12 months.
- There is a difference in the autocorrelation function (ACF) of the time series for rail breaks regarding the two asset types track and S&C. Both ACF indicate a seasonal pattern. However, while the ACF for track has both negative and positive autocorrelation, the ACF for S&C only has positive autocorrelation. The reason for the pattern in the ACF for rail breaks in S&C indicates that the time series is not stationary, while it is stationary for track. The time series plots of rail breaks in S&C and track respectively support this assumption. Hence, the increasing trend for all rail breaks can mainly be attributed to S&C, which indicates that the population of S&C is deteriorating over time.
- The CCF between NDT-inspections and rail breaks is about twice as large for track (0.4) as for S&C (0.2). This indicates that the NDT-inspections have more influence on the rail breaks for track than for S&C. This is probably due to the difference in maintenance practice between linear assets and point assets. Hence, machine inspections are less valuable for point assets than for linear assets due to the geographical extension of the latter. In addition, the additional safety inspections that are included in the preventive maintenance may be more effective for S&C compared to track and thereby influence the cross correlation. Hence, potential rail breaks are detected and corrected before they develop into

actual rail breaks in the S&C safety inspections. This means that both the NDT-inspections and safety inspections affect the occurrence of actual rail breaks in S&C and thereby reduce the cross correlation between the two individual preventive maintenance efforts and rail breaks.

- Three track sections with different characteristics suitable for further cause-and-effect studies have been identified. These are track section 124 (major preventive maintenance through NDT-inspections), 417 (few NDT-inspections, but other kinds of preventive maintenance), and 401 (large impact from corrective maintenance).
- There are NDT-remarks that are not rectified within the stipulated time. Further analysis is necessary to conclude whether these remarks represents an increased risk for actual rail breaks or an opportunity to improve the preventive maintenance practice. Here it would be interesting to benchmark with the practice of other infrastructure managers (e.g. Network Rail in England) and their experiences, as well as those that are responsible for running the NDT-train (i.e. Sperry).

In this study the amount of traffic and the occurrence of potential and actual rail breaks were linked on a track section level for the studied time period. To get a more appropriate cause-effect relationship it is necessary to identify at which specific track the rail break has occurred and the amount of traffic that it has experiences since installation. This approach would also support a more stringent reliability analysis. However, to succeed with this approach it is necessary to collect some additional data about characteristics of the infrastructure from the asset register (BIS).

One further suggestion for further research is to include the costs of delay to quantify the consequences of rail breaks, also considering derailments. Further work could also include a more thorough study of the corrective maintenance efforts, to highlight improvement possibilities regarding the maintenance support performance, e.g. regarding administrative and logistic delay times.

Acknowledgments The financial and intellectual support received from Trafikverket (Swedish transport administration) is gratefully acknowledged. Especially the contributions from Arne Nissen and Malin Syk are appreciated.

References

1. Cannon DF, Edel KO, Grassie SL, Sawley K (2003) Rail defects: an overview. Fatigue Fract Eng Mater Struct 26(10):865–886
2. Podofillini L, Zio E, Vatn J (2006) Risk-informed optimization of railway tracks inspection and maintenance procedures. Reliab Eng Syst Saf 91(1):20–35
3. Zhao J, Chan AHC, Stirling AB (2006) Risk analysis of derailment induced by rail breaks: a probabilistic approach. In: Proceedings of the annual reliability and maintainability symposium, RAMS 06—2006, 23–26 Jan 2006, pp 486–491

4. Kumar S, Espling U, Kumar U (2008) Holistic procedure for rail maintenance in Sweden. Proc Inst Mech Eng Part F J Rail Rapid Transit 222(4):331–344
5. Kumar S (2008) Reliability analysis and cost modeling of degrading systems. Doctoral thesis. Luleå University of Technology
6. Kumar U, Chattopadhyay G, Larsson-Kråik P-O (2005) Study of NDT rail inspection on Malmbanan. Research report, Banverket, Luleå
7. Yin RK (2003) Case study research: design and methods. Sage, Thousand Oaks
8. Montgomery DC, Jennings CL, Kulahci M (2008) Introduction to time series analysis and forecasting. Wiley, Hoboken
9. Rigdon SE, Basu AP (2000) Statistical methods for the reliability of repairable systems. Wiley, New York
10. Johnson RA, Wichern DW (2007) Applied multivariate statistical analysis. Prentice-Hall, New Yersey
11. ISO/IEC (2009) 31010:2009 risk management—risk assessment techniques. International Electrotechnical Commission, Geneva, Switzerland
12. TRV (2008) Säkerhetsbesiktning av fasta järnvägsanläggningar BVF 807.2—(Safety inspections of railway infrastructure assets) (2005). Trafikverket, Borlänge (In Swedish)

An Integrated Approach to Remaining Life Prediction of Rotating Machines

Tarun Chugh, V. Sankaranarayanan and P.V. Varde

Abstract State-of-the-art decisions related to repair and replacement of rotating machines in any process or industrial environment is still based on qualitative engineering judgment which often tends to be arbitrary and conservative in nature and has potential for loss of revenue and more importantly net production. The subject acquires significant dimension as number of rotating machines, for example, induction motors are very high, say of the order of 10–1000 or more for a plant like nuclear or process plant. Thus, there is a need to have a science or rational based approach for the plant managers to take decisions related to maintenance or ageing assessment based on well defined quantitative metrics or criteria. Industry experience suggests that prediction of health of insulation in an induction motor is one of the important parameters that require attention towards characterizing the life of the machine. This paper presents R&D work being performed on predicting the remaining useful life of insulation of the induction motors. The focus of this R&D is on development of an integrated framework where data driven approach is integrated to physics-of—failure approach towards developing robust model for predicting the remaining life of insulation.

Keywords Induction motors · Residual useful life (RUL) · Physics-of-failure · Winding insulation

T. Chugh (✉) · V. Sankaranarayanan
Nuclear Power Corporation of India Limited, Anushakti Nagar,
Mumbai 400094, India
e-mail: tarun.chugh91@gmail.com; tchugh@npcil.co.in

V. Sankaranarayanan
e-mail: vsankaranarayanan@npcil.co.in

P.V. Varde
Bhabha Atomic Research Centre, Mumbai 400085, India
e-mail: varde@barc.gov.in

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_39

1 Introduction

Rotating machines like induction motors are complex electro-mechanical devices utilized in most industrial applications for the conversion of power from electrical to mechanical form. Although induction motors are constructed, tested, and qualified to rigorous standards, failures of electric motors in nuclear power plants continue to occur. Operating anomalies, failures of other equipment, and other unforeseen circumstances can all contribute to aging degradation in motors. Recent studies regarding the operating experience of electric motors and the effects of aging on electrical equipment in nuclear power plants have indicated that many electric motor failures can be attributed to the aging and degradation of insulating materials and bearings caused by high temperature, vibration, moisture and other stressors.

Healthiness of the machines contributes to the production, down time reduction, reliability and revenues. Monitoring of the healthiness of the machines, therefore, is very important and essential. The ability to accurately predict changes in properties/parameters of electrical machines is of critical importance in optimizing the maintenance schedule of the plant. Thus, there is a continuous need to device test methods or to find more searching and sensitive parameters to predict the machine health. In view of this, it becomes quite important for a maintenance engineer to be able to predict the health of induction motor leading to appropriate usage of the machine(s), reduction in downtime, enhanced operational reliability and safety and revenues. Thus, the maintenance action can be optimized by diagnostics and prognostics methods which form a part of Condition Based Maintenance (CBM).

The stressors that affect large electric motors are: Heat, Chemicals, Pressure, Steam, Radiation, Mechanical Cycling/Rubbing, Humidity/Water Spray, Electromagnetic Cycling, Vibration/Seismic, and Foreign Object Ingestion.

The stressors act independently and/or synergistically to cause failures in the major subcomponents of large electric motors, such as the stator windings, electrical terminations, bearings, and rotor cage. All of the stressors listed above contribute to the gradual or catastrophic degradation of the insulation system. Mechanical and electromagnetic cycling, ingestion of foreign objects, and vibration-related stressors act upon the mechanical integrity of the machine. They can cause bearing and lubrication system problems, rotor breakage, mounting/enclosure failures, and failures of the shaft/couplings.

The stator winding system plays an important role in induction motors. A well designed stator winding insulation system can prevent the electrical short. There are several components and features in a stator winding insulation system, such as strand (sub-conductor) insulation, turn insulation and ground-wall (or ground or earth) insulation. Turn insulation is used for preventing shorts among the turns in the coil. The electrical insulation system decides the lifetime of insulation. For different purposes of insulation in electrical equipment, there are several [1].

In this paper, various approaches to study the insulation degradation phenomenon in induction motors like Failure Mode Effect Analysis (FMEA), Fault Tree Analysis (FTA) and Fuzzy Logic are reviewed and analyzed.

2 FMEA and FTA for Insulation Degradation

In order to understand the relationships of the various stressors to large motor operational performance, a failure modes and effects analysis (FMEA) was performed as shown in Table 1. The FMEA provides a systematic procedure for determining how each component of a device or system can fail, the mechanisms that cause it to fail, and how it can affect the overall performance of the device or system. The means for detection of the identified failure mechanisms are established along with methods for mitigating the effects of the failure mechanisms [2].

Fault Tree Analysis of Insulation Breakdown

One of the most critical component of an induction motor and also one of the main sources of their failure is the stator winding insulation system. Various

Table 1 FMEA for induction motor

Component name	Failure mode	Failure mechanisms	Failure effects
Stator winding	Winding to ground fault	Thermal degradation of insulation due to high ambient temperature, restricted ventilation, under- or over-voltages, low frequency, mechanical overload, voltage imbalance, single-phasing, too frequent starting, high process fluid temp, dust or dirt accumulation	Electrical trip
		Mechanical degradation of insulation due to vibration and rubbing	Damage to motor winding requiring
		Breakdown of insulation due to electrical transients and surges	Rewind
		Degradation of insulation due to moisture, lubricant, chemical reactions, or dirt	
		Manufacturing defect in insulation	
		Mechanical damage from loose part or ingested part	
	Winding-to winding fault	Same as above	Same as above
	Turn-to-turn fault	Same as above	Same as above
	Open winding	Breakdown of insulation and melting of conductors due to electrical transients and surges	Same as above for failure
		Broken winding conductor due to vibration, electromagnetic transients, and/or cyclic fatigue	

(continued)

Table 1 (continued)

Component name	Failure mode	Failure mechanisms	Failure effects
Stator leads and coil cross-ties	Phase-to-ground fault	Same as above	Same as above
	Phase-to-phase Fault	Same as above	Same as above
	Open circuit	Breakdown of insulation and melting of conductors due to electrical transients and surges	Same as above
		Broken conductor due to vibration, electromagnetic transients, and/or cyclic fatigue	
		Mechanical damage from loose part or ingested part	
		Mechanical damage from contact with rotating part	
	Loose leads or coil cross-ties	Loosening of leads, coil crossties, and fasteners due to vibration, electromagnetic transients, and/or cyclic fatigue	Degradation and damage to insulation and conductors
		Mechanical damage from loose part or ingested part	
Stator core	Loose laminations and locking bars in stator core assembly	Loosening of stator core assembly due to vibration	Increased losses (heat) due to larger leakage flux
		Loosening of stator core assembly due to electromagnetic transients	Increased motor current
		Misalignment of core assembly during manufacture	
	Lamination overloading	Thermal degradation and wear of lamination insulation	Increased losses (heat) due to excessive current in iron core
			Increased motor current

(continued)

Table 1 (continued)

Component name	Failure mode	Failure mechanisms	Failure effects
Rotor squirrel cage assembly	Rotor bars cracked at end ring	Fatigue due to vibration and mechanical cycling	Increased rotor cage resistance and heating
		Fatigue due to electro-magnetic cycling and transients	Increased vibration and wear of core laminations insulation
		Defective welds or brazed joints	Crack adjacent bars due to increased flexure
	Rotor bars loose in core slots	Loosening due to vibration and mechanical cycling	Increased vibration and wear of core laminations insulation
		Loosening due to electromagnetic cycling and transients	
		Loosening due to thermal cycling and excessive starting	Same as above
		Defective swaging during manufacture	
	Broken rotor bar	Same as above	Same as above
Rotor core	Loose laminations and locking bars in stator core assembly	Loosening of stator core assembly due to vibration	Increased losses (heat) due to larger leakage flux
		Loosening of stator core assembly due to electromagnetic transients	Increased motor current
		Misalignment of core assembly during manufacture	
	Lamination overheating	Thermal degradation and wear of lamination insulation	Increased losses (heat) due to excessive current in iron core.
			Increased motor current

surveys on motor reliability have been carried out over the years where the percentage of motor failures due to problem with the insulation is about 26 %. The unscheduled process downtime caused by a failure of the insulation system can cause enormous costs. FTA for stator winding insulation failure of induction motor is shown in Fig. 1.

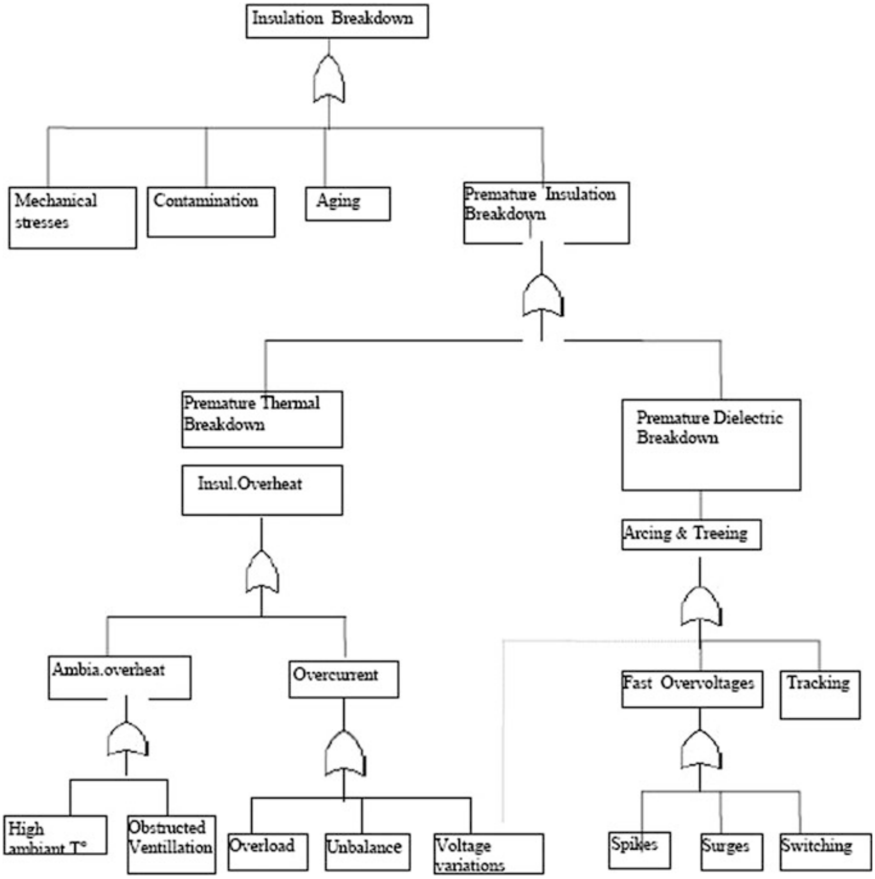


Fig. 1 FTA of insulation failure in a motor

3 Fuzzy Logic for RUL Estimation

In this approach, the d-q model for an induction motor is simulated using MATLAB/SIMULINK. The d-q model requires that all the three-phase variables have to be transformed to the two-phase synchronously rotating frame. Consequently, the induction machine model has blocks transforming the three-phase voltages to the d-q frame and the d-q currents back to three-phase [3].

The induction machine model implemented in this paper is shown in Fig. 2.

Simulation of Insulation Degradation [4]

To study the effect of stator insulation degradation in an induction motor, it is assumed that a shunt resistance is added in series with the phase impedance of the stator and the value of this resistance can be changed to vary the effect of degradation of stator insulation. The insulation sheets between the slots and coils and on

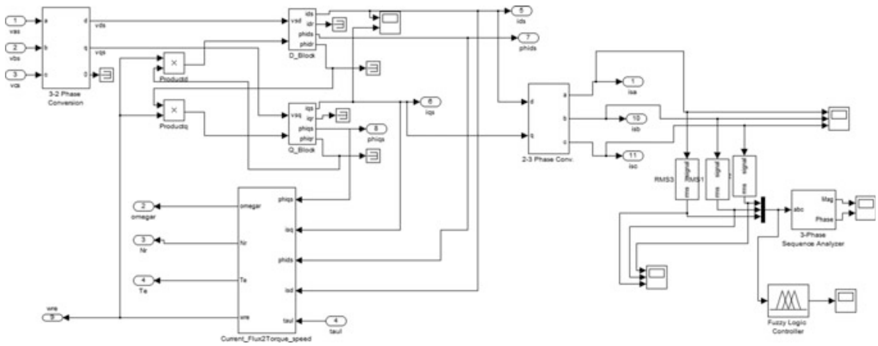


Fig. 2 Simulink d-q model of an induction motor

the enameled wires and between the turns in the coil are consists of different classes of insulation material. The choice depends on the maximum temperature rise permissible for each class. For each class of insulation material there exist limiting temperatures beyond which deterioration sets in and progresses rapidly. The degradation of insulation of the respective phase results in the reduction of resistance to the thermal conductivity. This results into higher heat transfer from the surface of the stator windings to the remaining part of structure. The interpretation of higher heat transfer is that the motor is subjected to extra load. Hence, motor draws more current as compared to what it would have drawn had the insulation not degraded. The increase in current is interpreted as reduction in the steady-state equivalent resistance of motor referred to stator. This is based on the interpretation of the presence of shunt resistance (Rsh) distributed across the entire phase as represented in the Fig. 3. The degree of degradation of, stator insulation is estimated on the basis of unbalance in stator phase currents.

For the induction motor of particular make, type and the frame size of suitable rating employed for particular application in an industry, it is probable that the status of degradation of stator insulation varies over considerable finite range. The finite range extends from certain minimum to permissible maximum value. The variation in the range of degradation particularly depends on several factors such as the time period over which motors are in use and the operating conditions.

The induction motor of particular make, type and the frame-size of suitable rating employed for particular application in an industry usually operate under variable load condition. In general, the motor is operating with the presence of many recipient faults, which are observable at corresponding frequencies in stator current spectrum. These include the stator winding faults due to several causes like inter-turn short circuit etc.; but other than the fault of stator insulation degradation. The fault with the stator insulation degradation would appear at supply frequency. In the spectrum the amplitude of current corresponding to supply frequency has two components; one due to actual load and other due to the stator insulation degradation. In order to fetch the current component due to stator insulation degradation it is therefore essential to eliminate the current component due to load. Hence, it is

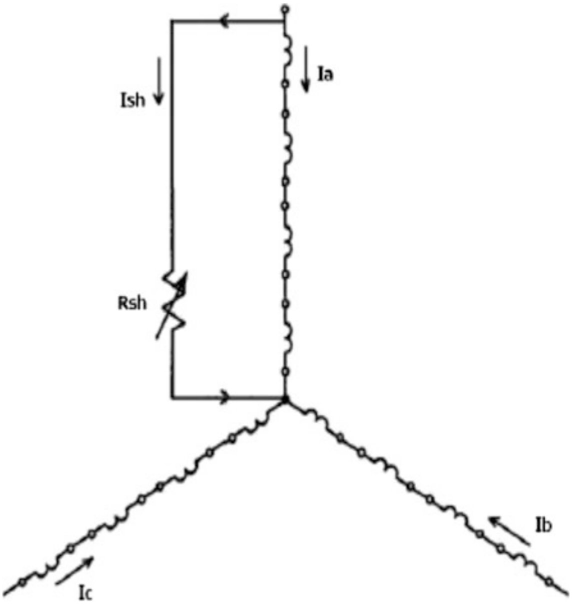


Fig. 3 Insulation degradation in an induction motor

obvious that the motor is to be operated at minimum mechanical load to ascertain the degree of degradation of stator insulation.

As a function of mains phase variables (ia, ib, ic) the Current Parks vector components (id, iq) are:

$$id = \sqrt{2/3}\,ia - \frac{ib}{\sqrt{6}} - \frac{ic}{\sqrt{6}} \tag{1}$$

$$iq = \frac{ib}{\sqrt{2}} - \frac{ic}{\sqrt{2}} \tag{2}$$

Under ideal conditions, three phase currents lead to a Current Park vector with the following components.

$$id = \sqrt{6/2} * I\,sinwt \tag{3}$$

$$id = \sqrt{6/2} * I\,sin\left(wt - \frac{\pi}{2}\right) \tag{4}$$

Where;

I Maximum value of the supply phase current;
w Supply frequency;
t Time variable

The percentage of degradation of stator insulation per phase is determined on the basis of unbalance in stator phase currents.

Ideal condition refers to zero-percent (0 %) state of percentage of degradation of stator insulation of the respective phases. First, the respective three-phase stator currents viz. Ia, Ib and Ic in amps; are computed and then on the basis of three to two-phase transformation model its Current Park vector pattern is determined. The Current Park vector representation is a circular pattern centered at the origin of the coordinate-axis as illustrated by Fig. 4.

When the stator insulation is degraded, then the plot between id and iq is an ellipse. The area under the ellipse increases with increase in degradation of the stator winding. The simulation results implies that each one of the current park vector data pattern is unique in representing the degree of unbalance in three-phase stator current on account of specific state of degradation of stator insulation of respective phases. This is shown in the Fig. 5 [5].

Fuzzy Logic [6–8]

A stator current signal contains potential fault information. The most suitable measurements for diagnosing the faults under consideration, in term of easy accessibility, reliability, and sensitivity, are the stator current amplitudes Ia, Ib, and Ic.

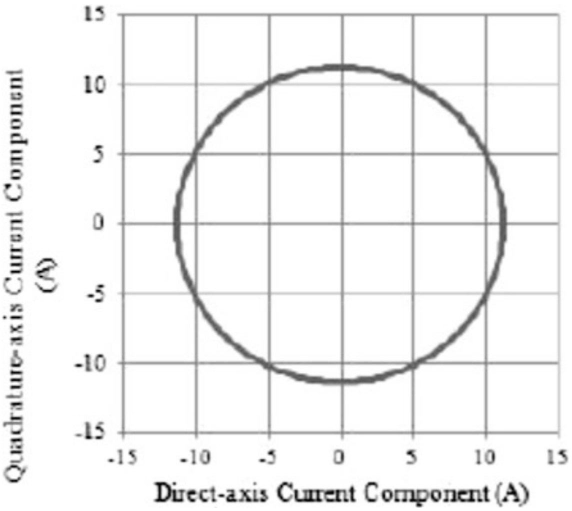
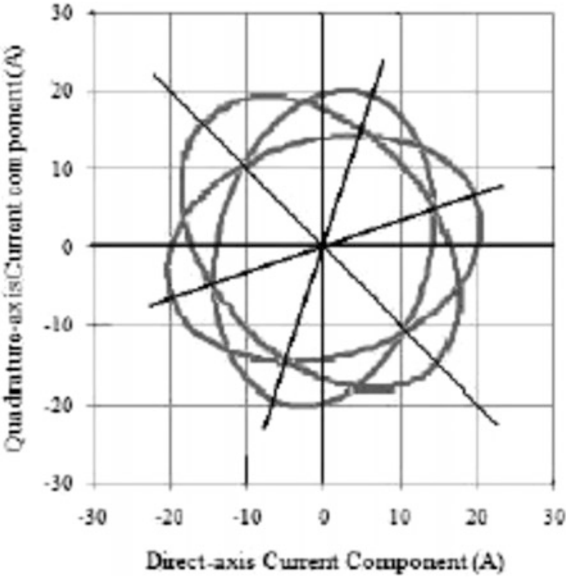


Fig. 4 Id-Iq—induction motor (balanced condition)

Fig. 5 Id-Iq induction motor (un-balanced condition)



Fuzzy systems rely on a set of rules. These rules, while superficially similar, allow the input to be fuzzy, i.e. more like the natural way that humans express knowledge. Thus, a power engineer might refer to an electrical machine as "somewhat secure" or a "little overloaded". This linguistic input can be expressed directly by a fuzzy system. Therefore, the natural format greatly eases the interface between the engineer knowledge and the domain expert. Furthermore, infinite graduations of truth are allowed, a characteristic that accurately mirrors the real world, where decisions are seldom "crisp".

As stated, the induction motor condition can be deduced by observing the stator current amplitudes. Interpretation of results is difficult as relationships between the motor condition and the current amplitudes are vague. Therefore, using fuzzy logic, numerical data are represented as linguistic information.

In our case, root mean square (rms) the stator current amplitudes Ia, Ib, and Ic are considered as the input variables to the fuzzy system. The stator condition in terms of the Residual Useful Life (RUL) is chosen as the output variable. All the system inputs and outputs are defined using fuzzy set theory.

$$Ia = \left\{ \frac{\mu ia(Iaj)}{Iaj} \in Ia \right\} \tag{5}$$

$$Ib = \left\{ \frac{\mu ib(Ibj)}{Ibj} \in Ib \right\} \tag{6}$$

$$Ic = \left\{ \frac{\mu ic(Icj)}{Icj} \in Ic \right\} \tag{7}$$

$$RUL = \left\{ \frac{\mu rul(RULj)}{RULj} \in RUL \right\} \tag{8}$$

where Iaj Ibj, Icj and RULj are, respectively, the elements of the discrete universe of discourse Ia, Ib, Ic, and RUL. µia(Iaj) µib(Ibj) µic(Icj) and µrul(RULj) are, respectively, the corresponding membership functions.

Basic tools of fuzzy logic are linguistic variables. Their values are words or sentences in a natural or artificial language, providing a means of systematic manipulation of vague and imprecise concepts. More specifically, a linguistic variable is characterized by a quintuple (x, T(x), U, G, M), where x is the variable name; T(x) is the set of names of the linguistic values of x, each a fuzzy variable, denoted generically by x and ranging over a universe of discourse U. G is a syntactic rule for generating the names of x values; M is the semantic rule associating a meaning with each value.

For instance, the term set T (RUL), interpreting stator condition, RUL, as a linguistic variable, could be

$$\mathrm{T(RUL)} = \{100\%,\ 80\%,\ 60\%,\ 40\%,\ 20\%\} \tag{9}$$

where each term in T (RUL) is characterized by a fuzzy subset, in a universe of discourse RUL. Good might be interpreted as a stator with no faults, damaged as a stator with voltage unbalance, and seriously damaged as a stator with an open phase.

Similarly, the input variables Ia, Ib, and Ic are interpreted as linguistic variables, with

$$\mathrm{T(Q)} = \{\mathrm{Zero,\ Small,\ Medium,\ Big}\} \tag{10}$$

where Q = Ia, Ib, Ic, respectively.

Fuzzy rules and membership functions are constructed by observing the data set. For the measurements related to the stator currents, more insight into the data is needed, so membership functions will be generated for zero, small, medium, and big. For the measurement related to the stator condition, it is only necessary to know if the stator condition is good, damaged, or seriously damaged. Once the form of the initial membership functions has been determined, the fuzzy if-then rules can be derived. Figure 6 depicts the linguistic variables used in this study. Membership functions for the input and output functions are shown in Figs. 7 and 8 respectively.

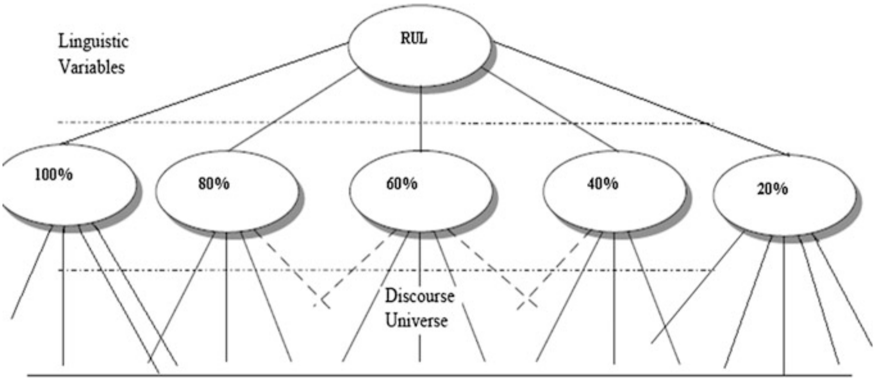


Fig. 6 Fuzzy logic—linguistic variables

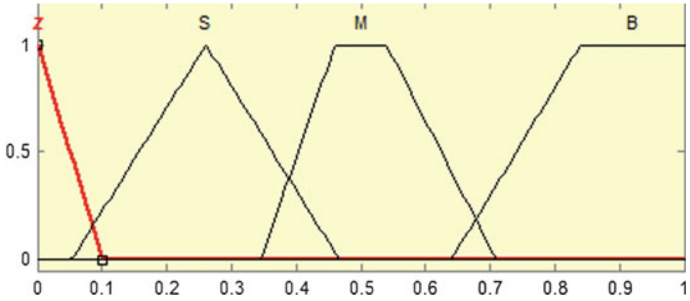


Fig. 7 Membership function of Ia, Ib and Ic

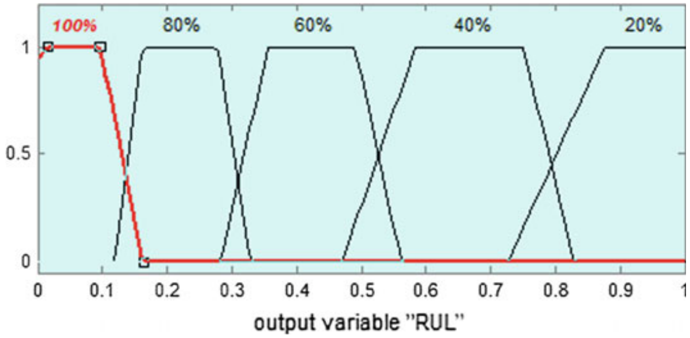


Fig. 8 Membership function for RUL

These rules have been optimized so as to cover all the healthy and the faulty cases. For ourstudy, we have obtained the following 14 *if-then* rules.

Rule (1): If *Ia* is *Z* Then *RUL* is *20 %*
Rule (2): If *Ib* is *Z* Then *RUL* is *20 %*
Rule (3): If *Ic* is *Z* Then *RUL* is *20 %*
Rule (4): If *Ia* is *B* Then *RUL* is *20 %*
Rule (5): If *Ib* is *B* Then *RUL* is *20 %*
Rule (6): If *Ic* is *B* Then *RUL* is *20 %*
Rule (7): If *Ia* is S and *Ib* is S and *Ic* is *M* Then *RUL* is *60 %*
Rule (8): If *Ia* is *S* and *Ib* is *M* and *Ic* is *M* Then *RUL* is *60 %*
Rule (9): If *Ia* is *M* and *Ib* is *S* and *Ic* is *M* Then *RUL* is *60 %*
Rule (10): If *Ia* is *M* and *Ib* is *M* and *Ic* is *M* Then *RUL* is *100 %*
Rule (11): If *Ia* is *S* and *Ib* is *S* and *Ic* is *S* Then *RUL* is *100 %*
Rule (12): If *Ia* is *S* and *Ib* is *M* and *Ic* is *S* Then *RUL* is *60 %*
Rule (13): If *Ia* is *M* and *Ib* is *S* and *Ic* is *S* Then *RUL* is *60 %*
Rule (14): If *Ia* is *M* and *Ib* is *M* and *Ic* is *S* Then *RUL* is *60 %*

Figure 9 shows the relationship of Ia with RUL as established by the Fuzzy Rule.

The simulation is run for two cases. In the first case, the stator insulation is healthy and the stator current and the derived id and iq are shown in Figs. 10 and 11 respectively. The Fuzzy Controller output is low signifying healthy condition as shown in Fig. 12.

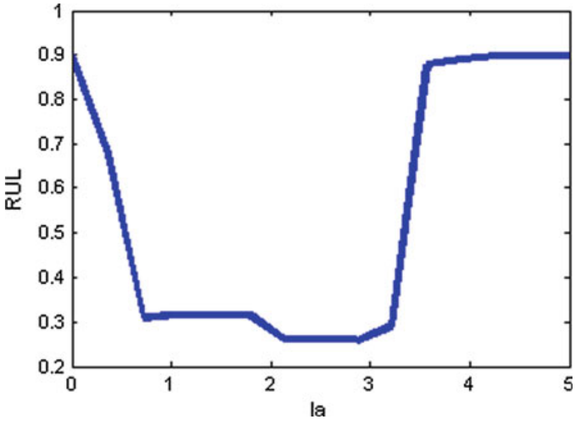


Fig. 9 Relationship of Ia with RUL

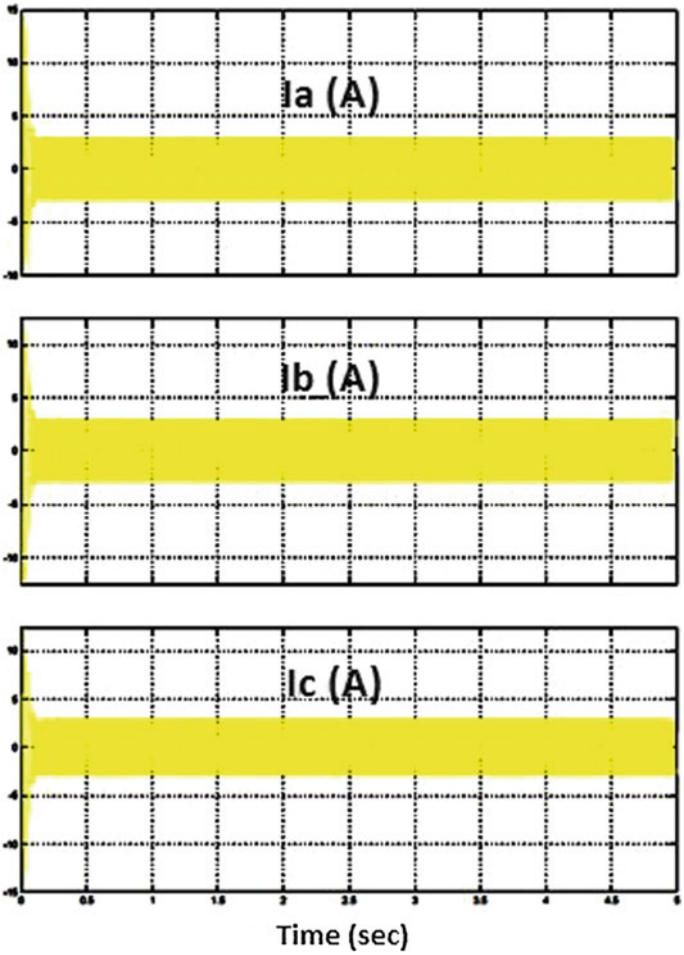


Fig. 10 Ia, Ib and Ic in healthy condition of insulation

Fig. 11 Id and Iq in healthy condition of insulation

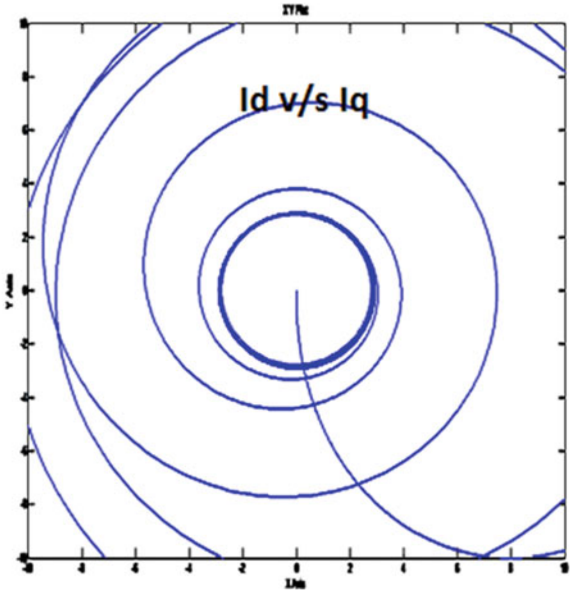
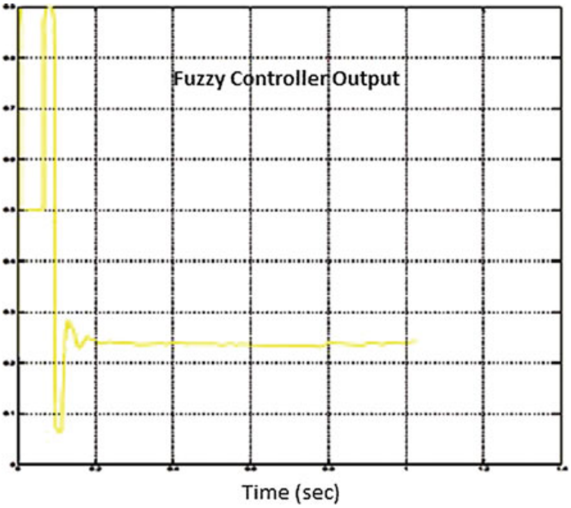


Fig. 12 Fuzzy controller output in healthy condition of insulation



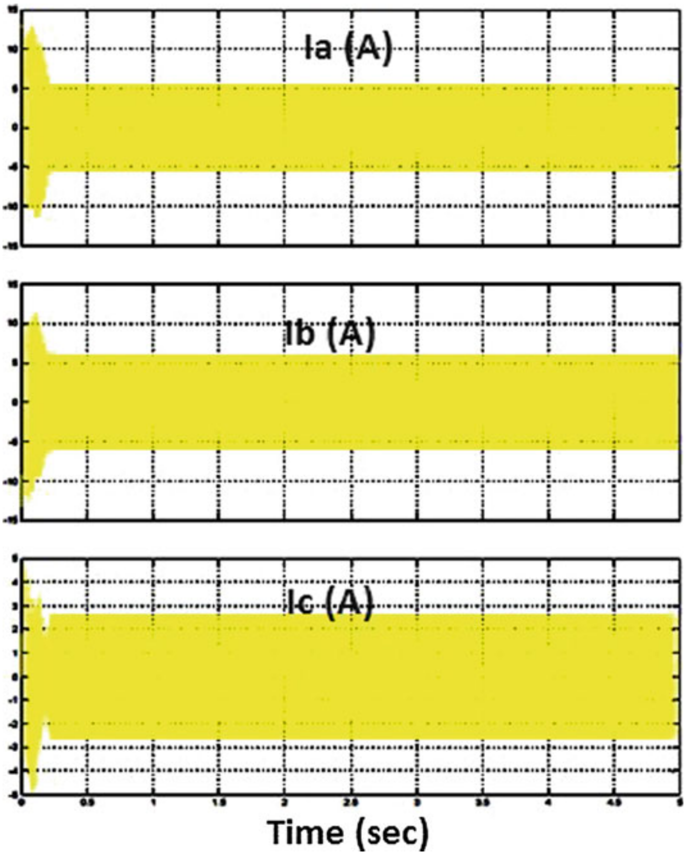


Fig. 13 Ia, Ib and Ic in deteriorated condition of insulation

In Case 2, the insulation failure is assumed, thus the effective stator resistance is varied for phase C. This is shown in Fig. 13. The plot between id and iq is not a circle but an ellipse which is shown in Fig. 14. Finally, the Fuzzy Controller Output is high signifying deteriorated condition as shown in Fig. 15.

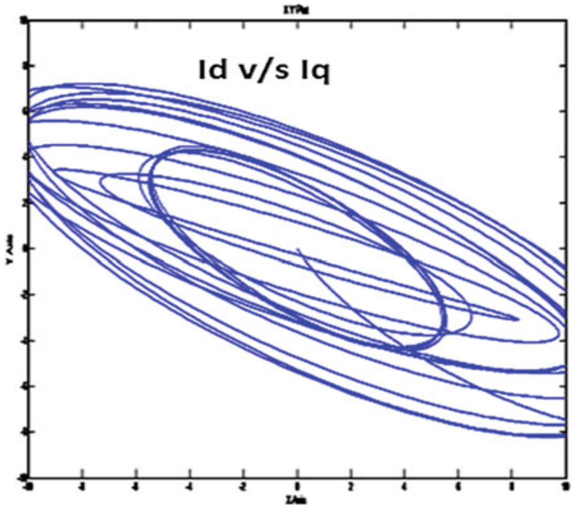


Fig. 14 Id and Iq in deteriorated condition of insulation

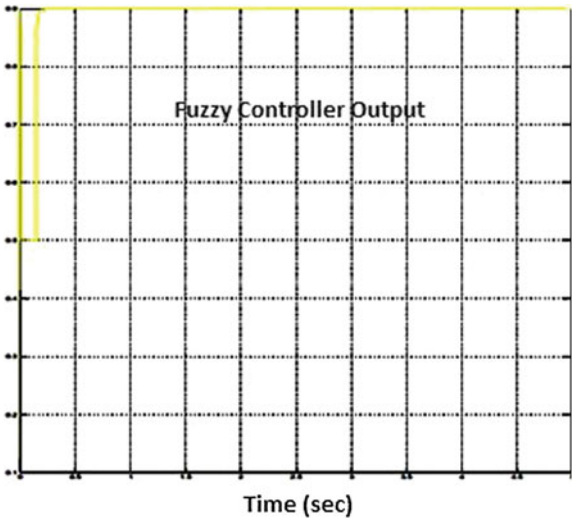


Fig. 15 Fuzzy controller output in deteriorated condition of insulation

4 Conclusions

In order to understand the relationships of the various stressors to large motor operational performance, a failure modes and effects analysis (FMEA) and Fault Tree Analysis (FTA) was performed. To simulate the effect of insulation degradation in an induction motor by variation of motor parameters, fuzzy logic has been used which estimate the RUL of the stator winding based in the stator currents. The Fuzzy Rule has thus been established. This helps estimation of the health of insulation based on a predefined set of rules. Additionally, the plot between id and iq is obtained which is also a signature analysis to predict the insulation health. Negative Sequence Current is also monitored.

In the present simulation, the induction motor has been modeled using its d-q model and RUL is estimated based on the stator current and graph between id and iq.

Acknowledgments Our thanks to sponsors of ICRESH-ARMS 2015 for their intellectual and financial support.

References

1. Pillay P, Manyage M (2006) Loss of life in induction machines operating with unbalanced supplies. IEEE transactions on energy conversion 21(4)
2. Vfflaran M, Subudhi M (1996) Aging assessment of large electric motors in nuclear power plants. Brookhaven National Laboratory
3. Ozpineci B, Tolbert LM (2003) Simulink Implementation of induction machine model—a modular approach. International Conference on Electric Machines and Drives 2:728–734
4. Modak AJ, Inamdar HP (2010) Computer-aided simulation model of stator ground-wall insulation of induction motor based on current Park's vector approach. Int J Comput App 9 (8):0975–8887
5. Anbarasu E, Karthikeyan M (2013) Modeling of induction motor and fault analysis. Int J Eng Sci Innovative Technol (IJESIT) 2(4)
6. Zeraoulia M, Mamoune A, Mangel H, Benbouzid MEH (2005) A simple fuzzy logic approach for induction motors stator condition monitoring. J Electrical Systems 1(1):15–25
7. Rodríguez PVJ, Arkkio A (2008) Detection of stator winding fault in induction motor using fuzzy logic. Appl Soft Comput 8(2):1112–1120
8. SaravanaKumar R, Vinoth Kumar K, Ray KK (2009) Fuzzy logic based fault detection in induction machines using lab view. Int J Comput Sci Netw Secur 9(9):226–243

Reliability Compliance Testing of Electronic Systems Using Parametric and Non Parametric Sequential Test Plans

Diana Denice, Manoj Kumar and P.P. Marathe

Abstract During design & development, reliability of electronic systems is predicted using well-known prediction methods like empirical models and life testing. Empirical models help to quantify reliability during design phase of systems while life testing or accelerated life testing methods are applied once the system is developed. Assessment of system reliability using these methods follows bottom-up approach and as a result, any latent uncertainty at the component/subsystem level gets amplified at the system level. To deal with the limitations of existing methods and to verify that system reliability goals are met, reliability compliance testing is gaining importance. In compliance testing, system reliability is not predicted but it is demonstrated by testing, whether system conforms to the system requirement or not. This method does not require any additional setup or chamber and it is based on the failure data of system components obtained during testing of systems after installation. Hence, it is very cost effective. In this paper, two sequential compliance test plans for control & instrumentation (C&I) system of a Nuclear Power Plant (NPP) are discussed, using parametric and nonparametric analysis. Parametric analysis assumes exponential time-to-failure distribution while nonparametric analysis is based on distribution free sequential rank-sum probability ratio test. A case study of these plans for a NPP C&I is also presented. Finally, a comparison of both the methods is made.

Keywords Reliability prediction · Compliance tests · Parametric estimation · Non parametric estimation

D. Denice (✉) · M. Kumar · P.P. Marathe
Control Instrumentation Division, BARC, Mumbai, India
e-mail: dianad@barc.gov.in

M. Kumar
e-mail: kmanoj@barc.gov.in

P.P. Marathe
e-mail: ppm@barc.gov.in

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_40

1 Introduction

For industrial as well as critical applications, reliability has grown to be one of the most significant attributes of electronic system design. During design & development, reliability is embedded into a system by choosing good quality components and raw materials, employing proven manufacturing technologies, redundancy etc. and improved by means of failure analysis and testing. However, once the system is ready, it becomes imperative to confirm the target reliability. In this case, reliability prediction helps to assess if system reliability goals are met.

There are two well-known methods in prediction i.e. empirical failure models and life testing. Empirical models are used during design phase of systems to predict initial reliability figures. While life testing method applied after system development gives estimates close to field reliability [1]. Also, accelerated life testing which is a modified life testing method works on the principle of accelerating failure mechanisms of components to estimate mean time to failure (MTTF).

However, assessment of system reliability using these methods has an inherent limitation, i.e. it follows bottom-up approach and as a result, any latent uncertainty at the component/subsystem level gets amplified at the system level. To deal with these shortcomings and to verify system reliability goals, reliability compliance testing seems promising.

In compliance testing, system reliability is not predicted but it is demonstrated by testing, whether system conforms to the system requirement or not. This method does not require any additional setup or chamber and uses failure data available during testing of systems after installation. Hence, it is very cost effective.

There are two basic types of compliance tests; sequential tests and time/failure terminated tests [2]. If all the samples are tested at a time, it is called a time/failure terminated test or fixed sample test. If samples are tested sequentially, one-by-one or batch-by-batch it is called a sequential test. Sequential tests offer several benefits over their fixed counterpart. Few of them are (i) they require smaller test duration for very reliable or very unreliable items [2]. (ii) sample size required is smaller in many cases [3] which reduce the total cost of testing.

Therefore, in this paper, two types of sequential test plans: parametric and non parametric are discussed. Parametric sequential tests assume a parametric distribution to derive the plan while non parametric tests, as the name suggests are distribution free. In this paper, a probability ratio test is used in the parametric case and a rank-sum probability ratio test for non parametric case. A detailed plan is developed using both methods for C&I system of NPP. Finally, a case study of NPP C&I available in parametric analysis is applied to non parametric case and results are compared.

2 Sequential Analysis

Sequential analysis is a hypothesis testing situation in which the course of action is reassessed as observations become available [4]. As opposed to conventional hypothesis testing where the number of observations (i.e. sample size) is treated as a constant, sample size in sequential hypothesis testing depends on the outcome of observations. As a result, it is not fixed.

The sequential method of testing a hypothesis *H* is described as follows [5]. A rule is formulated for making one of the following three decisions at every observation of the experiment:

(i) Accept *H*
(ii) Reject *H*
(iii) Continue the experiment by taking additional observations.

If the first observation leads to acceptance or rejection of the test, the test is terminated. If no decision is reached, the test is continued with the second observation and so on.

2.1 Methods and Applications

Methods: There are two widely accepted methods in sequential testing: parametric methods and non-parametric methods [3].

In parametric methods, failure distributions like Gaussian, exponential, binomial etc. are assumed to derive the sequential test plan. Well established and most widely used procedure in this category is known as the Sequential Probability Ratio Test (SPRT).

To deal with those samples whose failure distribution is not known a priori, non parametric tests such as Sequential Signed Rank Test (SSRT), Wilcoxon Sequential Signed Rank Test (WSSRT) and Sequential Rank-Sum Probability Ratio Test (SRSPRT) are available.

Applications:

1. Sequential experimentation is tremendously used in the field of medical and pharmaceutical research. With the help of this method, only a few patients are tested (sequentially) for the efficacy of a new drug [6], instead of testing all the patients.
2. Another field of its application is in lot acceptance tests, especially when the items to be tested are very expensive or get destroyed during testing. It is recently standardised for use in reliability compliance testing [2].

In this paper, SPRT based on the likelihood ratio of sample and SRSPRT based on likelihood ratio of Wilcoxon signed rank statistic [7] are used. Out of the three non parametric tests available, SRSPRT is used because of its similarity to SPRT, which will get clearer from the following section.

2.2 Sequential Probability Ratio Test (SPRT)

Let us consider the simple null hypothesis,

$$H_0 : \theta = \theta_0$$

against the alternate hypothesis

$$H_1 : \theta = \theta_1$$

where θ is the parameter of interest.

Let $f(x, \theta)$ denote the distribution of a random variable x with θ as its parameter. Then, the distribution of x is given by $f(x, \theta_0)$ when H_0 is true and $f(x, \theta_1)$ when H_1 is true.

Successive observations on x are denoted by $x_1, x_2, x_3 \ldots\ldots x_n$.

For any positive integer m, the joint probability distribution of obtaining sample $x_1, x_2 \ldots\ldots x_m$ is given by [5]

$$p_{0,m} = f(x_1, \theta_0) * \ldots * f(x_m, \theta_0) \text{ when } H_0 \text{ is true} \tag{1a}$$

$$p_{1,m} = f(x_1, \theta_1) * \ldots * f(x_m, \theta_1) \text{ when } H_1 \text{ is true} \tag{1b}$$

where $x_1, x_2 \ldots\ldots x_m$ are independent and identically distributed (*i.i.d*) observations.

Two types of errors arise in any hypothesis testing [4]. Type I error also known as producer's risk α is the probability of rejecting H_0 when it is true i.e.

$$\alpha = P(H_1|H_0) \tag{2}$$

where $P(H_i|H_j)$ is the probability of accepting H_i when H_j is true.

Type II error also known as consumer's risk β is the probability of accepting H_0 when it is not true.

$$\beta = P(H_0|H_1) \tag{3}$$

Thus, SPRT for testing H_0 against H_1 is derived as follows:

For each observation x_m, compute the likelihood or probability ratio ($p_{1,m}/p_{0,m}$).

If

$$B < \frac{p_{1,m}}{p_{0,m}} < A \tag{4}$$

test is continued by taking next observation.

If

$$\frac{p_{1,m}}{p_{0,m}} \geq A \tag{5}$$

test is terminated with the rejection of H_0 (acceptance of H_1).

If

$$\frac{p_{1,m}}{p_{0,m}} \leq B \tag{6}$$

the test is terminated with acceptance of H_0.

The two constants A and B given [5] by

$$A = \frac{(1-\beta)}{\alpha}; \quad B = \frac{\beta}{(1-\alpha)} \tag{7}$$

which define the boundary conditions for the probability ratio.

2.3 Sequential Rank-Sum Probability Ratio Test (SRSPRT)

Consider the same hypotheses described in Sect. 2.1.

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta = \theta_1$$

Let x_1, x_2, x_3…… x_n be i.i.d observations and θ be their true mean.

SRSPRT makes the only assumption that the underlying distribution for samples is symmetric about its median, which is much easier to satisfy and more likely to be true than to assume a specific distribution function [7].

This test is based upon the probability ratio of Wilcoxon signed rank statistic W_i^+ under each hypothesis H_i; where i = 0, 1.

Then, the Wilcoxon signed rank statistic is given by

$$W_i^+ = \sum_{k=1}^{n} R\left(x_i^k\right)\psi\left(x_i^k\right) \tag{8}$$

where W_i^+ is a discrete random variable, x_i^k is the *k*th ordered data in the sample and $R(x_i^k)$ is rank of x_i^k, *i* = *0, 1* is the hypothesis index and *j* = *1, 2,.....n* is sample index.

To obtain the Wilcoxon statistic under hypothesis $H_{i,}$ compute

$$x_{j,i} = x_j - \theta_i \tag{9}$$

Now arrange $|x_{j,i}|$ in increasing order of magnitude. Assign a rank *R* according to their position and an indicator ψ according to their sign.

$R(x_i^k) = K$ is the rank of the data $|x_{j,i}|$

$$\psi(x_i^k) = \begin{Bmatrix} 1 & if\ x_i^k > 0 \\ 0 & if\ x_i^k < 0 \end{Bmatrix} \tag{10}$$

Then, the rank-sum probability ratio [7] is given by

$$r_n = \frac{P(W_1^+, n)}{P(W_0^+, n)} \tag{11}$$

If

$$B < r_n < A \tag{12a}$$

test is continued by taking additional observation.

If

$$r_n \geq A \tag{12b}$$

test is terminated with the rejection of H_0 (acceptance of H_1).

If

$$r_n \leq B \tag{12c}$$

the test is terminated with acceptance of H_0.

2.4 Operating Characteristic Curve (OCC)

Once the hypothesis is formulated and a test plan is developed, it is best to know the performance of the same plan or probability of acceptance for all the other hypotheses which can arise from the formulated hypothesis i.e. OC curve depicts the probability of acceptance of hypothesis under test not only when $\theta = \theta_0$ and $\theta = \theta_1$ but also when $\theta \neq \theta_0$ and $\theta \neq \theta_1$.

It is a curve which plots the probability of acceptance of H_0,$L(\theta)$ against the parameter θ under test. OC curve for a sequential plan is derived from equation [5]

$$L(\theta) \sim \frac{A^h - 1}{A^h - B^h} \tag{13a}$$

where h is the root of the equation

$$\int_{-\infty}^{+\infty} \left[\frac{f(x,\theta_1)}{f(x,\theta_0)}\right]^h f(x,\theta)dx = 1 \tag{13b}$$

Solving the above integral w.r. to h by substituting $f(x,\theta)$ gives an expression for h. It is then substituted into (11) along with A and B to get $L(\theta)$.

3 Development of Test Plans

Sequential test planning involves stating the requirement in the form of hypothesis and fixing the sampling risks. Sample size is not fixed in advance. Then, using one of the sequential procedures discussed above, a sequential plot and test truncation criteria are worked out.

An Integrated Test Facility (ITF) has been set up at BARC. The purpose of ITF is to validate the C&I systems of an NPP to their requirements. C&I consist of 22 racks and 80 types of modules like power supply, processors, signal conditioning, I/O hardware etc. The total quantity of modules is around 1000 nos.

The time to failure of modules during the test will be used to check the compliance to reliability requirement. The reliability requirement for the entire C&I system is Mean Time To Failure (MTTF) of 1000 h.

Two test plans (SPRT and SRSPRT) are developed for the same requirement. The aim is to compare the two and bring out their merits and demerits.

Hypothesis for testing MTTF θ of C&I system is proposed as

$$H_0\colon \theta \geq 1000\,\text{h}$$
$$H_1\colon \theta \leq 500\,\text{h}$$

3.1 Using SPRT

Time to failure of the entire C&I system is assumed to be exponentially distributed with mean θ. Hence, the probability of failing r times in an accumulated test time t is given by the Poisson distribution [2],

$$p(r)=\left(\frac{t}{\theta}\right)^{r}\exp\left(-\frac{t}{\theta}\right)\left(\frac{1}{r!}\right) \tag{14}$$

The sequential probability ratio becomes

$$\frac{p_1(r)}{p_0(r)}=\left(\frac{\theta_0}{\theta_1}\right)^{r}\exp\left(-\left(\frac{1}{\theta_1}-\frac{1}{\theta_0}\right)t\right)$$

Substituting it into Eq. (10),

$$B<\left(\frac{\theta_0}{\theta_1}\right)^{r}\exp\left(-\left(\frac{1}{\theta_1}-\frac{1}{\theta_0}\right)t\right)<A$$

Substituting *A, B* from (7) and taking natural logarithm,

$$\frac{\ln(B)}{\ln\left(\frac{\theta_0}{\theta_1}\right)}+\frac{\left(\frac{1}{\theta_1}-\frac{1}{\theta_0}\right)}{\ln\left(\frac{\theta_0}{\theta_1}\right)}t<r<\frac{\ln(A)}{\ln\left(\frac{\theta_0}{\theta_1}\right)}+\frac{\left(\frac{1}{\theta_1}-\frac{1}{\theta_0}\right)}{\ln\left(\frac{\theta_0}{\theta_1}\right)} \tag{15}$$

Now, assuming α = 10 %, β = 10 %, the following equation is obtained

$$-3.17+1.443\frac{t}{\theta_0}<r<3.17+1.443\frac{t}{\theta_0}$$

Accept line: $r=-3.17+1.443\frac{t}{\theta_0}$

Reject line: $r=3.17+1.443\frac{t}{\theta_0}$

The next step is to determine the truncation criteria for the test. For an exponential distribution, the test truncation time t_t is calculated using the formula [2],

$$t_t=\frac{\theta_0\chi^2_{\alpha,2r_t}}{2} \tag{16a}$$

where the test truncation failure number r_t is determined from the ratio [7]

$$\frac{\chi^2_{\alpha,2r_t}}{\chi^2_{1-\beta,2r_t}}\geq\frac{\theta_1}{\theta_0} \tag{16b}$$

Using the above formulae, test truncation ratio $\left(\frac{t_t}{\theta_0}\right)$ is found to be 9.883 and truncation failure number r_t to be 15.

Since, the sample size *n* is not predetermined in advance, as per the truncation criteria, test will be truncated when *n* = 15 samples have failed or data points fall into acceptance or rejection region for *n* < 15.

Sequential testing plot with truncation is shown in Fig. 1.

Fig. 1 Sequential SPRT plot

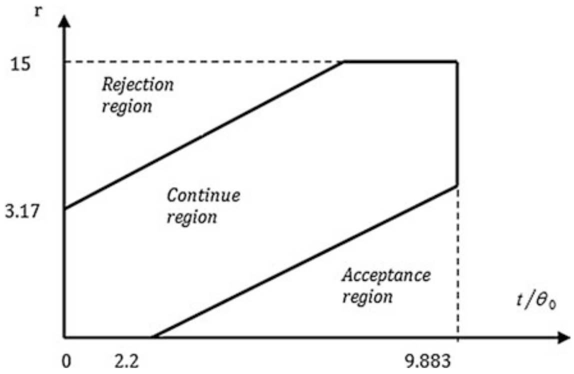
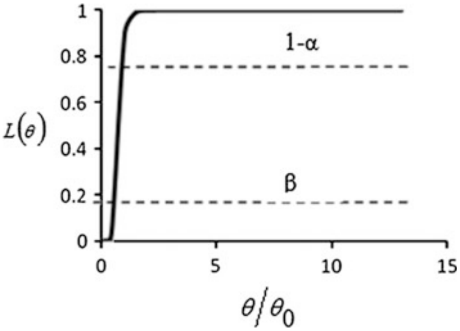


Fig. 2 OC plot



From the plot, it is seen that a failure free operation of 2200 h will lead to acceptance of H_0 while more than 3 failures within 2200 h will lead to rejection.

The OC function for this test at ITF is calculated from the pair of equations given below and plotted as shown in Fig. 2. To compute θ, values of h (positive and negative) are chosen such that it takes a range of values lesser than θ_1 and higher than θ_0. At the same time, $L(\theta)$ is computed for the same values of h; now the curve is plotted with θ on x-axis and $L(\theta)$ on y-axis [7]

$$L(\theta) = \frac{A^h - 1}{B^h - A^h} \tag{17a}$$

$$\theta = \theta_0 \frac{\left(\frac{\theta_0}{\theta_1}\right)^h - 1}{h\left(\frac{\theta_0}{\theta_1} - 1\right)} \tag{17b}$$

In the OC plot, when θ is 1000 h the probability of acceptance of H_0 is 89.99 % ($1-\alpha$) and when θ is 500 h, the probability of acceptance of H_0 is only 9 % (β).

As can be seen, this sequential plan discriminates well between H_0 and H_1.

3.2 Using SRSPRT

For sample size, $n \geq 10$, W_i^+ can be approximated by a Gaussian distribution with mean $\bar{W} = \frac{n(n+1)}{4}$ and variance $V_T = \frac{n(n+1)(2n+1)}{24}$[8].

i.e. the pdf of W_i^+ becomes

$$P\left(W_i^+, n\right) = \frac{1}{\sqrt{2\pi V_T}} \exp\left(-\frac{\left(W_i^+ - \bar{W}\right)^2}{2V_T}\right) \tag{18}$$

Substituting (14) in (7) and taking natural logarithm on both sides gives the ratio r_n as

$$\ln r_n = \frac{12}{n(n+1)(2n+1)}\left(W_0^+ - W_1^+\right)\left(W_0^+ + W_1^+ - \frac{n(n+1)}{2}\right) \tag{19}$$

Substituting it in $\ln B < \ln r_n < \ln A$

We get

$$\frac{n(n+1)(2n+1)}{12}\ln B < \left(W_0^+ - W_1^+\right)\left(W_0^+ + W_1^+ - \frac{n(n+1)}{2}\right) < \frac{n(n+1)(2n+1)}{12}\ln A$$

Let

$$y = \left(W_0^+ - W_1^+\right)\left(W_0^+ + W_1^+ - \frac{n(n+1)}{2}\right) \tag{20}$$

then

$$\frac{n(n+1)(2n+1)}{12}\ln B < y < \frac{n(n+1)(2n+1)}{12}\ln A \tag{21}$$

Accept line: $y = \frac{n(n+1)(2n+1)}{12}\ln B$

Reject line: $y = \frac{n(n+1)(2n+1)}{12}\ln A$

Since a minimum sample size of 10 is assumed in the derivation, test will be truncated for $n = 10$.

Now, assuming $\alpha = 10\ \%$, $\beta = 10\ \%$, $n = 1$ to 10, the following plot is obtained as shown in Fig. 3.

From this graph, it can be seen that at least one failure must occur to reach a decision unlike SPRT which takes care of a no failure case.

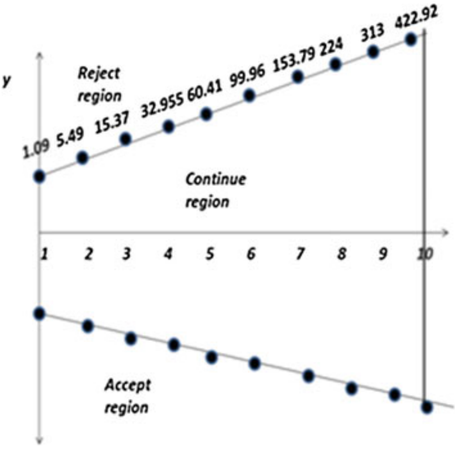


Fig. 3 Sequential SRSPRT plot

4 Case Study

A truncated parametric sequential test plan available for entire C&I of NPP is taken up as a case study to understand the planning and execution of sequential analysis. The sequential test plan is as follows.

Given data: $\theta_0 \geq 252\,\text{hrs}$; $\theta_1 \leq 100\,\text{hrs}$; α = 25 %; β = 25 %, where all these notations retain the same meaning as above.

Using this data, accept and reject line equations are obtained using (13a, 13b). Truncation criteria are obtained from (16a) and (16b). i.e.

Accept line: $r = -1.188 + 1.645\frac{t}{\theta_0}$;

Reject line: $r = 1.188 + 1.645\frac{t}{\theta_0}$

Truncation criteria:$r_t = 2.5$; $t_t/\theta_0 = 1.34$

If the system meets this criterion, it is said to be compliant, otherwise non-compliant.

Data was collected when the entire C&I system was functional in its operating conditions as shown in Table 1.

Plotting the failure data in sequential plot in Fig. 4,

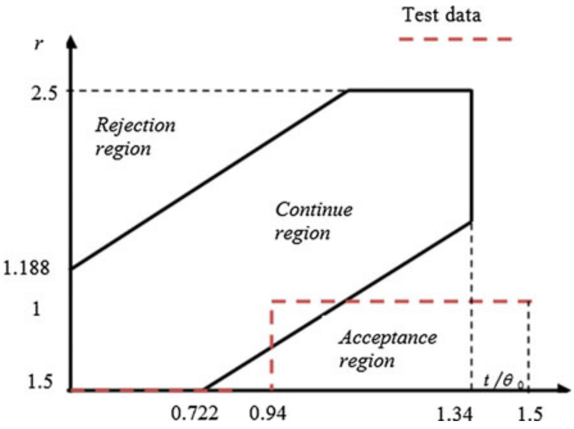
From the plot, it is seen that

$$r = 1; t/\theta_0 = 1.5$$

Table 1 Data collected during the test on C&I

r	t (hrs)	t/θ_0
1	237	0.94
2	461	0.94–2.52

Fig. 4 Sequential SPRT plot showing experimental data



Since,

$$r < r_t;\ t/\theta_0 > t_t/\theta_0$$

Hence, it is demonstrated that C&I of NPP is compliant with the reliability requirement.

5 Comparison of Sprt and Srsprt

In order to compare the results of SPRT with SRSPRT, data obtained in the case study is applied to SRPRT plan using the same α, β and MTTF values.

Data is transformed using (8), (9), (10) and (20) and shown in Table 2:

Accept and Reject line equations are given by (21) as

$$y = \frac{-1.0986n(n+1)(2n+1)}{12};y = \frac{1.0986n(n+1)(2n+1)}{12}$$

which are symmetric about n.

Truncation criteria: n = 10

The sequential plot with the data is shown in Fig. 5.

Table 2 Data collected in case study applied to SRSPRT

n	t (hrs)	$t - \theta_0$	$t - \theta_1$	W_0^+	W_1^+	y
1	237	−15	137	0	1	0
2	461	209	361	2	3	−2

Fig. 5 Sequential SRSPRT plot showing experimental data

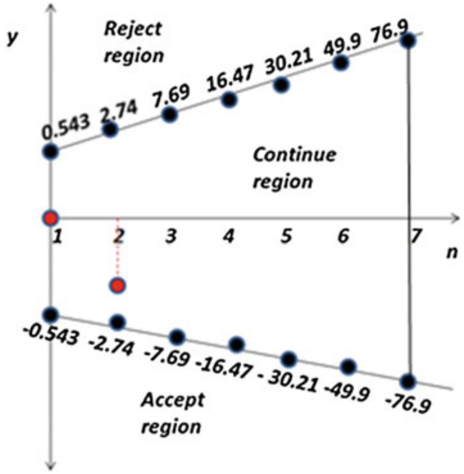


Table 3 SPRT versus SRSPRT

Attribute	SPRT	SRSPRT
Distribution	Requires prior knowledge of the distribution	Does not require prior knowledge of distribution
Robustness of the plan	Because of dependency on a model, it is not robust if actual model is different from assumed one.	Robust as it doesn't make any model assumptions
Sample size	Requires smaller sample size for the same test strength (α, β)	Higher sample size for the same test strength (α, β)

From the plot, it is seen that 2 observations are not sufficient to make any decision about the hypotheses under test. This plan needs to be supported by more data points to reach a conclusion.

Based on the theory and case study, an attribute comparison matrix is made for these methods as discussed in Table 3.

6 Conclusion

Reliability compliance testing of C&I systems using sequential analysis is presented here. In this test, reliability is not predicted but demonstrated, only to verify the system requirement. The test can be conducted during testing of systems after installation and hence does not require any additional setup or chamber. Sequential plans are preferred over conventional fixed sampling plans because in most of the cases, they require smaller test duration and sample size. It is preferable to go for an SPRT plan when the sampling distribution is known a priori because it requires smaller sample size compared to SRSPRT. A detailed step-by-step planning and procedure is given for both SPRT and SRSPRT, which can be applied to any type of C&I systems, for any underlying distribution, once the requirements are known.

References

1. Denice D, Kumar M (2012) Life testing of electronic modules by multiple accelerated life tests. In: Proceedings of ICQRIT-ITM, Delhi
2. IEC (2006) 61124: Reliability testing-Compliance tests for constant failure rate and constant failure intensity. In: International electrotechnical commission, Geneva, Switzerland
3. Ghosh BK, Sen PK (1991) Handbook of sequential analysis. Marcel Dekker Inc, New York
4. Kapur KC, Lamberson LR (1977) Reliability in engineering design, 1st edn. Wiley Inc, New York
5. Wald Abraham (1963) Sequential analysis, 6th edn. John Wiley & Sons Inc, New York
6. Denice D, Kumar M, Marathe PP (2014) Reliability testing of C&I systems using sequential test plans. In: Proceedings of SACI, Mumbai
7. Chenggang Yu, Bingjing Su (2004) A non–parametric sequential rank-sum probability ratio test method for binary hypothesis testing. J Signal Process 84:1267–1272
8. Miller RG (1997) A sequential signed rank test. Technical report. Stanford University at California, California

Bayesian Reliability with MCMC: Opportunities and Challenges

Jing Lin

Abstract The recent proliferation of Markov Chain Monte Carlo (MCMC) approaches has led to the use of the Bayesian inference in a wide variety of fields, including reliability engineering. With the current (and future) proliferation of new products, old problems continue to hamper us, while new challenges keep appearing. In Bayesian reliability, these include but are not limited to: (1) achieving and making use of prior information; (2) applying small data sets or system operating/environmental (SOE) data with big and complex data; and (3) making posterior inferences from high-dimensional numerical integration. To deal with old problems while meeting new challenges, this paper proposes an improved procedure for Bayesian reliability inference with MCMC, discussing modern reliability data and noting some applications where the Bayesian reliability approach with MCMC can be used. It also explores opportunities to use Bayesian reliability models to create stronger statistical methods from prior to posterior. Finally, it outlines some practical concerns and remaining challenges for future research.

Keywords Reliability · Bayesian statistics · Markov chain monte carlo (MCMC) · System operating/environmental (SOE) data · Big reliability data

1 Introduction

The Bayesian framework is more attractive now that it is comparatively easy for users to incorporate what they know about the world into their conclusions and to calculate how probabilities change as new evidence arises [1]. Bayesian reliability offers modern methods and techniques for analysing reliability data from a Bayesian perspective [2, 3]. It has been popular with reliability engineers for over

J. Lin (✉)
Division of Operation, and Maintenance Engineering, Luleå University of Technology, 97187 Luleå, Sweden
e-mail: janet.lin@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_41

50 years because of its ability to consider information from past experiments, impressions, prejudices, etc. [4], especially as it is generally time-consuming (or cost consuming) to obtain enough failure data for a reliability assessment. However, its application has been restricted because of the difficulties calculating high-dimensional numerical integration from posterior information.

Markov Chain Monte Carlo (MCMC) is essentially Monte Carlo integration using Markov chains. It draws samples from the required distribution and forms sample averages to approximate expectations. With these samples, it runs a cleverly constructed Markov chain for a long time. The proliferation of MCMC approaches has led to the use of the Bayesian inference in a wide variety of fields for the past two decades [2, 5].

In Bayesian reliability, the advances in MCMC approach have opened up new possibilities, with research appearing in book chapters and research papers. Work has been done on the following topics and their cross-applications (discussed in a later section): (1) hierarchical reliability models; (2) fault tree analysis; (3) complex system reliability analysis; (4) change points analysis; (5) accelerated failure models; (6) masked system reliability; (7) degradation analysis; (8) accelerated degradation testing; (9) deterioration analysis; (10) life cycle reliability assessment; (11) updating of structural models and reliability; (12) reliability of repairable system; (13) software reliability models, etc. To implement modern computational-based Bayesian approaches for reliability inference, Lin (2014) proposes a general approach for Bayesian reliability using MCMC methods, developing a procedure consisting of four stages and 11 steps [6].

Bayesian reliability will undoubtedly be popular in the future, but it is still criticized. First, most forms of prior distributions are motivated by mathematical tractability. This limits the practical usefulness of the models, as it does not facilitate the elicitation of valid engineering judgements in the form of a prior distribution. Second, although implementing MCMC to get posterior distributions can compensate the bias associated with difficulties of choosing a suitable prior, ongoing problems include posterior sampling, MCMC convergence diagnostic, Monte Carlo error diagnostic, etc. Furthermore, considering the ongoing proliferation of new products, the next generation of reliability data comprises System Operating/Environmental (SOE) data (with big, complex characters) [7, 8]. The question of how to deal with SOE data (or big reliability data) must be resolved to fully implement Bayesian approaches in reliability.

To deal with old problems while meeting new challenges and to inspire future research, this paper introduces some examples using Bayesian reliability with MCMC. It proposes an improved procedure for its implementation, and explores some of the opportunities to use Bayesian reliability approaches with MCMC. The rest of this paper is organised as follows. Section 2 presents some examples of Bayesian reliability applications with MCMC. Section 3 discusses modern reliability data, noting old problems and new challenges from prior to posterior; Sect. 4 proposes an improved procedure for Bayesian reliability inference with MCMC. Section 5 explores some of the opportunities and practical concerns in the use of

Bayesian reliability with MCMC, including stronger statistical methods from prior to posterior. Section 6 offers concluding remarks and outlines areas for further research.

2 Examples of Bayesian Reliability Applications with MCMC

Examples of Bayesian reliability with MCMC include but are not limited to the following topics and their cross-applications.

- A building block approach to model validation may proceed through various levels, such as material to component to subsystem to system, comparing model predictions with experimental observations at each level. In hierarchical reliability modelling, reliability data usually become scarce as one proceeds from lower to higher levels. In the Bayesian structural equation modelling method for hierarchical model validation, a Bayesian network with MCMC is applied to represent the two relationships and to estimate the influencing factors between them [9]. Graves et al. [10] discuss how to use simultaneous higher-level and partial lower-level data in reliability assessments.
- Using a genetic algorithm and MCMC methods, Hamada et al. [11] propose a fully Bayesian approach that simultaneously combines non-overlapping (in time) basic and higher-level event failure data in fault tree quantification.
- Lin [12] proposes a Bayesian reliability approach for detecting certain change-points, which may disturb the evaluation of reliability models with covariates, via a two-stage failure model and stochastic time-lagged regression functions.
- In masked system lifetime data, the exact component causing the system's failure is often unknown. Bayesian reliability modelling with Gibbs sampling and MCMC approaches have been proposed to model the masking probabilities [13].
- To solve the combined problem of small data samples and incomplete datasets whilst simultaneously considering the influence of several covariates, Lin et al. [14–17] apply both parameter and non-parameter Bayesian reliability models (incl. Bayesian survival analysis) with MCMC to the analysis of degradation of locomotive wheel-sets considering their different installed positions. Lin et al. [18] also propose a Bayesian reliability analysis with MCMC for complex systems, where a certain fraction of the subsystems is defined as a "cure fraction" under the consideration that such subsystems' lifetimes are long enough and, in fact, never fail during the life cycle of the entire system.
- Accelerated degradation testing (ADT) is a common approach in reliability, especially as it is time-consuming and cost-consuming to obtain enough field failure data. Wang et al. [19] propose a Bayesian reliability evaluation method

with MCMC to integrate the ADT data from the laboratory with the failure data from the field.

- Given the limited number of variables that can be controlled and observed, unobserved heterogeneity is almost inevitable in a reliability study. Most existing models do not fully account for the heterogeneity issue. Hong and Prozzi [20] adopt Bayesian reliability with MCMC approaches to make pavement deterioration forecasts at different confidence levels with varying inspection frequencies.
- Pan [21] proposes Bayesian reliability to improve product reliability prediction by integrating failure information from both the field performance data and the accelerated life testing data. Furthermore, as modern technology continues to advance, available data for reliability assessment of a new product are extremely sparse and sometimes contain subjective information. A Bayesian model updating approach has been developed to evaluate new products' life cycle reliability [22]. To better understand how to update sectional time-to-failure (TTF) distribution as new operational TTF data become available, Briand and Huzurbazar [23] propose a Bayesian change point methodology to combine lifecycle failure distribution.
- Bayesian reliability with MCMC has been applied for model updating, as some claim [24] there are always modelling errors associated with constructing a theoretical model of the behaviour of a structure, and this leads to uncertain accuracy in the predicted response.
- Relying on the use of MCMC methods, a Bayesian model which takes into account missing data is proposed to describe failures in a complex, expanding over time, repairable system, split into components installed over different years [25].
- Software reliability is one of the most significant attributes of software quality. To handle group data and time point data, Hirata et al. [26] propose a unified MCMC algorithm based on the Metropolis-Hasting method, regardless of data structures. Aktekin and Caglar [27] develop a Bayesian model with imperfect debugging in software reliability considering multiplicative failure rate. They use actual inter-failure data to carry out inference testing on model parameters via MCMC and present additional insights from Bayesian analysis.
- Bayesian reliability with MCMC has been applied in some comparison studies. For instance, Soliman et al. [28] investigate the problem of point and interval estimations for the modified Weibull distribution (MWD) using progressively type-II censored sample. The maximum likelihood (ML), Bayes, and parametric bootstrap methods are used for comparing estimations from the unknown parameters as well as some lifetime parameters (reliability and hazard functions). Based on a general, data-driven framework, Lin et al. [15] undertake a general reliability study using both classical and Bayesian semi-parametric degradation approaches to illustrate how to flexibly determine reliability to support preventive maintenance strategy making.

3 Big and Complex Reliability Data Meet Bayesian

3.1 Modern Reliability Data

Meeker and Hong (2014) call the next generation of reliability data System Operating/Environmental (SOE) data [7]. These data are characteristically "big" and "complex" [8]. To the typical three "Vs" (Volume, Velocity and Variety), the concept of big reliability data adds more "Vs", for instance, Veracity, Value, Visualization, Volatility, Validity, Venue, Vocabulary, and Vagueness. Complexity stems from high dimensionality, poor data quality, complex relationships, incomplete data and many other properties of big data, caused by the increasing use of numerous types of sensors, mobile device, tether-free, web-based applications and other information and communication technologies (ICT).

Along with the new technologies, new challenges keep appearing. In Bayesian reliability, these include but are not limited to: (1) achieving and making use of prior information; (2) applying both small data sets and system operating/environmental (SOE) data with big, complex characters; (3) making posterior inferences from high-dimensional numerical integration.

3.2 Prior

In Bayesian reliability, traditional prior knowledge comes from a wide range of historical information, including: engineering design, component test data, system test data, operational data from similar systems, field-tracking studies in various environments, computer simulations, related standard and operation manuals, experience data from similar systems, expert judgment and personal experience, warranty data, etc. It also takes a variety of forms, including reliability data, the distribution of reliability parameters, moments, confidence intervals, quantiles, and upper and lower limits.

In modern reliability data, the information which can be used as prior will embrace the attributes enumerated by the "Vs" mentioned above, together with "complexity". Not surprisingly, given the nature of "big prior information", there are challenges in the acquisition, inspection, fusion and selection of such knowledge for MCMC.

3.3 Model

The challenge in the next generation of reliability data includes creating high quality reliability models. The model must be robust, as customers think there is enough big prior knowledge to be used.

In addition to the various forms of traditional reliability parametric models, semi-parametric models, frailty models and other untraditional reliability models, advanced MCMC stochastic models, Bayesian time series models, Bayesian belief networks, Bayesian quantile regression models, Bayesian causal inference, etc. will receive increasing attention.

Bayesian reliability models with MCMC will become more popular in maintenance modelling, RAMS (reliability, availability, maintainability, safety), and CBM (condition based maintenance), etc. Already researchers are interested in how to use additional data or historical information to support decision making.

3.4 Posterior

Given the burgeoning interest, we need to develop more advanced posterior sampling approaches with today's big data. Posterior/model diagnostics and updated approaches are required.

In addition, the loss function must be studied further, as most reliability decision support models are actually based on the square loss function from the posterior results, and this is inadequate.

4 An Updated Procedure for Bayesian Reliability Reliability with MCMC

Lin (2014) proposes a general approach to Bayesian reliability using MCMC methods by developing a procedure consisting of four stages and 11 steps [6]. An improved procedure (see Fig. 1) is composed of a continuous improvement process and includes 16 sequential steps. By implementing the step-by-step procedure, we can accumulate and gradually update prior knowledge. Equally, posterior results will be improved upon and become increasingly robust, thereby improving the accuracy of the inference results. Details of the procedure include:

- Step 1: Data collection. The original data sets for prior information and current data related to target reliability studies must be identified and acquired. Various data sources are discussed in Sect. 3.2.
- Step 2: Data preparation. Collected prior information needs to be evaluated, cleaned, and merged. In this way, prior information can be transferred to prior knowledge, and current data can become data for reliability analysis in later steps.
- Step 3 and Step 4: Prior inspection and integration. In these steps, prior knowledge receives a second and more extensive treatment, including but not limited to: a reliability consistency check, a credence test, and a multi-source integration. These steps improve prior knowledge.

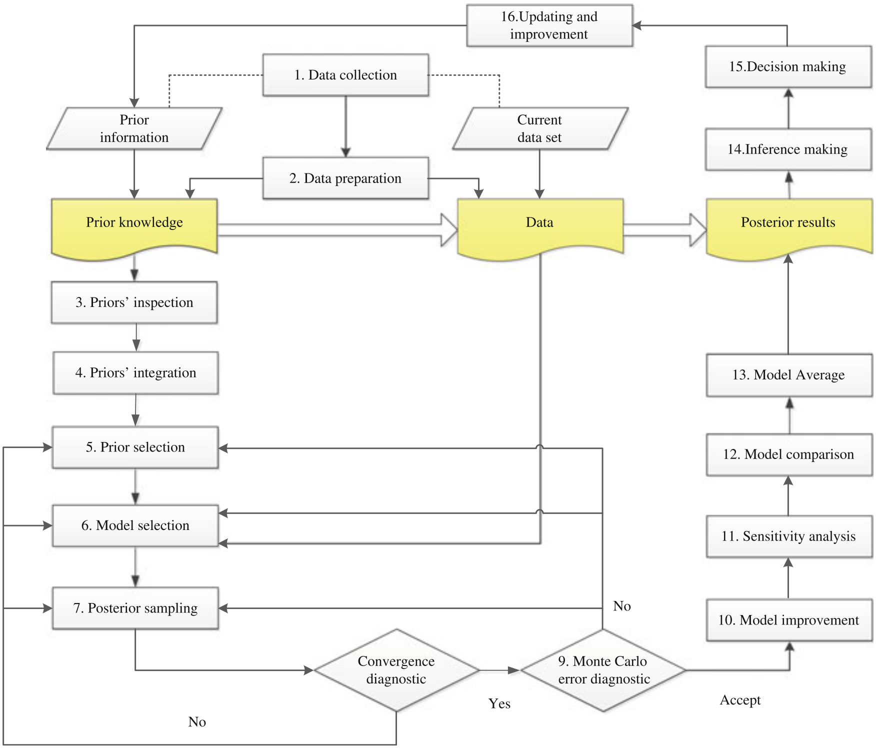


Fig. 1 A procedure for bayesian reliability inference via MCMC

- Step 5: Prior selection. This step determines the model's form and parameters.
- Step 6: Model selection. This step determines a reliability model (see Sect. 3.3), selecting i^{th} $(i = 1, \cdots i+1, \cdots n)$ model from n candidates for the studied system/units.
- Step 7: Posterior sampling. In this step, we determine a sampling method (for instance, Gibbs sampling, Metropolis-Hastings sampling, etc.) to implement MCMC simulation for the model's posterior calculations.
- Step 8: Convergence diagnostic. In this step, we check whether the Markov chains have reached convergence. If so, we move to the next step; if not, we return to Step 7 and re-determine the iteration times of posterior sampling or re-choose the sampling methods; if the results still cannot be satisfied, we return to Steps 5 and 6 and re-determine the prior selection and model selection.
- Step 9: Monte Carlo error diagnostic. We need to decide if the Monte Carlo error is small enough to be accepted in this step. As discussed in Step 8, if it is accepted, we go on to the next step; if it is not, we return to Step 7 and re-decide the iteration times of the posterior sampling or re-choose the sampling methods; if the results still cannot be accepted, we go back to Steps 5 and 6 and recalculate the prior selection and model selection.

- Step 10: Model improvement. Here, we choose the $i + 1^{th}$ candidate model and restart from Step 6.
- Step 11: Sensitivity analysis. After implementing n candidate models, sensitivity analysis is implemented with different values of the prior to study the robustness of the Bayesian method.
- Step 12: Model comparison. After implementing n candidate models, we need to compare the posterior results to determine the most suitable model.
- Step 13: Model average. For the accepted candidate models, we need to adopt the average posterior estimations (using the Bayesian model average or the MCMC model average) as the final results.
- Step 14: Inference making. After achieving the posterior results in Step 13, we can perform Bayesian reliability inference.
- Step 15: Decision making. According to the selected loss function, we can determine system (or unit) reliability, find the failure distribution, and optimise maintenance strategies, etc.
- Step 16: Data updating and inference improvement. Along with the passage of time, new "current data" can be obtained, relegating previous inference results to prior information. By updating reliability data and prior knowledge, and restarting at Step 1, we can improve the reliability inference.

5 Potential Applications and Practical Concerns of Modern Reliability Data with Bayesian Inference

To improve the decision-making process in modern reliability, data from various sources (e.g. product, production, maintenance, and business) must be collected, integrated, fused, and analysed to transform them from information into knowledge. Given the quickly developing areas of Information and Communication Technologies (ICT), new knowledge-driven (or data-driven) approaches in computational sciences and applied mathematics must be developed to support reliability strategies to predict, diagnose or make a prognosis of a complex system's behaviour. Another important focus is how to use the previous and current results to make prescriptions to support engineers on site.

Increasing attention is being paid to big data analytics to extract information, knowledge and wisdom from big data [29]. In the reliability field, big data have huge potential to enable sophisticated knowledge-driven decision-making and facilitate new ways to organise, learn and innovate. However, as model-driven decision-making is still important when failure mechanisms cannot be achieved, Bayesian reliability is a good way to combine model-driven approaches and knowledge-driven approaches.

Although big data are starting to handle both visible and invisible issues in reliability, the research gaps in its application remain large. Professor Judea Pearl, winner of the 2011 A.M. Turing Award, notes that "big data must go to causal" [30].

He also says many data scientists remain unconcerned about the critical distinction between statistical and causal inference. In the future, however, causal inference combined with Bayesian belief networks will a play key role in reliability, from prediction to prognosis and prescription.

Bayesian quantile regression models [31] will start attracting more researchers considering reliability applications. This method can work well with prior and posterior knowledge, even with the complexity of modern reliability data.

Although some researchers suggest big data represent the “next big thing in innovation”, “the fourth paradigm of science”, or “the next frontier for innovation, competition, and productivity”, small data (or incomplete, censored/truncated, or no data) still exist and continue to challenge asset management. Meanwhile, no matter whether data are big or small, some “old problems (incl. uncertainties on the parameter estimation, complexity and large-scale system, dependences between events)” must be studied further [32].

Considering the harsher limitations imposed on decision makers today (incl. resources, cost, humanity, environmental influences), integrated reliability decisions must be studied together in Multi-objective Optimization Problems (MOP) models. Besides handling traditional technologies, Bayesian Optimization Algorithm (BOA) and Bayesian Belief Networks (BNN) can deal with MOP in a more flexible way, as their capabilities include considering prior information.

6 Conclusion

Bayesian approaches with MCMC will become increasingly popular in reliability engineering. With this in mind, this paper presents some examples of Bayesian reliability applications with MCMC. Considering modern reliability data, it proposes an improved procedure consisting of 16 steps. It argues Bayesian reliability is able to meet the new challenges of big prior and big posterior data, with huge potential to enable more sophisticated knowledge-driven decision making and combine knowledge with model-driven models. In the near future, both causal inference and Bayesian quantile regression will play a pivotal role in reliability analysis, from prediction to prognosis and prescription. That said, a number of old problems (incl. uncertainties on the parameter estimation, complexity and large-scale system, dependences between events) must be studied further.

References

1. Nuzzo R (2014) P values, the ‘gold standard’ of statistical validity, are not as reliable as many scientists assume. Nature 506:150–152
2. Hamada MS, Wilson A, Reese CS, Martz H (2008) Bayesian reliability. Springer, New York
3. Singpurwalla ND (2006) Reliability and risk: a bayesian perspective. Wiley, Chichester

4. Tillman FA, Kuo W, Hwang C, Grosh DL (1982) Bayesian reliability & availability—a review. IEEE Trans Reliab R-31(4):362–372
5. Gilks W, Richardson S, Spiegelhalter D (1996) Markov chain monte carlo in practice. Chapman and Hall/CRC, London
6. Lin J (2014) An integrated procedure for bayesian reliability inference using MCMC. J Qual Reliab Eng 2014
7. Meeker WQ, Hong Y (2014) Reliability meets big data: opportunities and challenges. Qual Eng 26:102–116
8. Göb R (2014) Discussion of "Reliability Meets Big Data: Opportunities and Challenges". Qual Eng 26:121–126
9. Jiang X, Mahadevan S (2009) Bayesian structural equation modeling method for hierarchical model validation. Reliab Eng Syst Saf 94(4):796–809
10. Graves T, Hamada M, Klamann R, Koehler A, Martz H (2008) Using simultaneous higher-level and partial lower-level data in reliability assessments. Reliab Eng Syst Saf 93 (8):1273–1279
11. Hamada M, Martz H, Reese C, Graves T, Johnson V, Wilson A (2004) A fully Bayesian approach for combining multilevel failure information in fault tree quantification and optimal follow-on resource allocation. Reliab Eng Syst Saf 86(3):297–305
12. Lin J (2008) A two-stage failure model for bayesian change point analysis. IEEE Trans Reliab 57(2):388–393
13. Kuo L, Yang TY (2000) Bayesian reliability modeling for masked system lifetime data. Stat Probab Lett 47(3):229–241
14. Lin J, Asplund M, Parida A (2014) Reliability analysis for degradation of locomotive wheels using parametric bayesian approach. J Qual Reliab Eng Int 30(5):657–667
15. Lin J, Julio P, Asplund M (2014) Reliability analysis for preventive maintenance based on classical and bayesian semi-parametric degradation approaches using locomotive wheel-sets as a case study. J Reliab Eng Syst Saf 134:143–156
16. Lin J, Asplund M (2014) A comparison study for locomotive wheels' reliability assessment using the Weibull frailty model. J Eksploatacja i Niezawodnosc-Maint Reliab 16(2):276–287
17. Lin J, Asplund M (2015) Bayesian semi-parametric analysis for locomotive wheel degradation using gamma frailties. Inst Mech Eng Proc Part F J Rail Rapid Transit 229(3):237–247
18. Lin J, Nordenvaad ML, Zhu H (2011) Bayesian survival analysis in reliability for complex system with a cure fraction. Int J Perform Eng 7(2):109–120
19. Wang L, Pan R, Li X, Jiang T (2013) A Bayesian reliability evaluation method with integrated accelerated degradation testing and field information. Reliab Eng Syst Saf 112:38–47
20. Hong F, Prozzi JA (2006) Estimation of pavement performance deterioration using bayesian approach. J Infrastruct Syst 12(2):77–86
21. Pan R (2009) A Bayes approach to reliability prediction utilizing data from accelerated life tests and field failure observations. Qual Reliab Eng Int 25(2):229–240
22. Peng W, Huang H, Li Y, Zuo MJ, Xie M (2013) Life cycle reliability assessment of new products-A Bayesian model updating approach. Reliab Eng Syst Saf 112:109–119
23. Briand D, Huzurbazar A (2008) Bayesian reliability applications of a combined lifecycle failure distribution. Proc Inst Mech Eng Part O J Risk Reliab 222:713–720
24. Beck JL, Au S-K (2002) Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation. J Eng Mech 128:380–391
25. Pievatolo A, Ruggeri F (2004) Bayesian reliability analysis of complex repairable system. Appl Stoch Model Bus Ind 20(3):253–264
26. Hirata T, Okamura H, Dohi T (2009) A bayesian inference tool for NHPP-based software reliability assessment. Future Gener Inf Technol 5899(2009):225–236
27. Aktekin T, Caglar T (2013) Imperfect debugging in software reliability: a bayesian approach. Eur J Oper Res 227(1):112–121
28. Soliman AA, Abd-Ellah AH, Abou-Elheggag NA, Ahmed EA (2012) Modified Weibull model: a bayes study using MCMC approach based on progressive censoring data. Reliab Eng Syst Saf 100:48–57

29. Russom P (2011) Big data analytics. The Data Warehousing Institute (TDWI), Renton
30. http://www.bayesia.us/blog/why-big-data-must-go-causal
31. Yu K, Moyeed RA (2001) Bayesian quantile regression. Stat Probab Lett 54(4):437–447
32. Zio E (2009) Reliability engineering: old problems and new challenges. Reliab Eng Syst Saf 94:125–141

Root Cause Analysis in Support of Event Investigation

N.S. Joshi, Sachin Kumar and P.V. Varde

Abstract Reliability based methods are widely used for the safety assessment of plant system, structures and components. These methods provide a quantitative estimation of system reliability but do not provide insight into the failure mechanisms. Understanding the failure mechanisms is a must to avoid the recurrence of the events and enhancement of the system reliability. Root Cause Analysis (RCA) provides a tool for gaining detailed insights into the causes of failure of a component with particular attention to the identification of fault in component design, operation, surveillance, maintenance, training, procedures and policies which must be improved to prevent repetition of events. Dhruva is a 100 MWth research reactor located at Bhabha Atomic Research Centre, Mumbai. In this research reactor, each failure/malfunction in SSCs or human error, termed as an 'Event' is reported by plant Operations and discussed in plant safety committee. Depending upon the importance of the events to the safety of the plant, RCA team of experts in various disciplines analyzes the events. This paper discusses the methodologies adopted for performing RCA of different events in Dhruva reactor.

Keywords Root cause analysis · Low level events · Significant events

1 Introduction

The most commonly and widely used event investigation technique in most of the nuclear power plants, research reactors facilities and regulatory bodies is Root Cause Analysis (RCA) or Root Cause Investigation. Most of the NPPs follow event reporting

N.S. Joshi (✉) · S. Kumar · P.V. Varde
RRSD, Bhabha Atomic Research Centre, Mumbai 400 085, India
e-mail: nsjoshi@barc.gov.in

S. Kumar
e-mail: s_kumar@barc.gov.in

P.V. Varde
e-mail: varde@barc.gov.in

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_42

system which requires investigations, reporting of occurrences, implementation and follow up of corrective actions. The level of efforts expended should be based on the significance attached to the occurrence. Most of the events occurring in NPPs need only a scaled down efforts while a few occurrences related to safety of the plant need to be investigated using one or more formal analytical methods.

At Reactor Group in Bhabha Atomic Research Cenre (BARC), a Task Force is in place to analyze the events occurring at the Research Reactors. The RCA methodology uses several techniques and their prime objective is to find the underlying cause i.e. the root cause that if properly addressed and corrected would prevent recurrence of similar events in the future. At Dhruva, an event reporting system already exists. The events are reported either as 'Event Reports' or 'Significant Event Reports' depending upon the severity of the event to the safety of the plant. All the faults and events are classified as 'Near Miss Events' and 'Low Level Events' [1]. Most of these events have no safety significance and significant impact on safety performance. It has been proposed to perform RCA of some of the consequential events and significant events. Low Level Event or Near Miss Event may not be significant to plant safety. An accumulation of low level events in the same system or with similar patterns may indicate a lack or failure in a surveillance programme. Multiple low level events may be considered as precursor for significant events. Experience has shown that a relationship exists between those events affecting nuclear safety, performance, reliability, and individual events that have no significant impact on performance. Considering the safety significance of the events, RCA methodology has been classified as Level-1 and Level-2. A Level-1 RCA methodology is adopted to analyze Low Level Events and the analysis will be informal. However, while investigating a significant event or a complex problem, Level-2 methodology comprising of several analytical methods/models is employed. Typically the decision is based on real or probable potential consequences of the event.

2 RCA Background

It was observed that the all the current root cause analysis methods could be traced to three TRADITIONS. The first and oldest of these traditions is the Management Oversight and Risk Tree Process (MORT) developed in the early 1970s for the U.S. government. This process is used extensively by the U.S. Nuclear Regulatory Commission (NRC) for special inspections and incident investigations and many nuclear plants worldwide. The second tradition was the Human Performance Enhancement System (HPES). This process was developed by the Institute of Nuclear Power Operations (INPO) in the early 1980s initially for use in the U.S. nuclear plants and later extended to international application through the World Association of Nuclear Operators (WANO). The third tradition was the Assessment of Safety Significant Events Team process (ASSET). ASSET was developed by the International Atomic Energy Agency (IAEA) for use in evaluating incidents by Member States.

While each of the three traditions contains unique elements, but their philosophies are similar. Each one identifies what could be considered ‘direct causes’ and through further analysis determines ‘root causes’. Each recognizes the importance of understanding the sequence of events, specific working level causes and determines the related management system causes. As on today, there are several RCA methods available. Most commonly methods [2, 3] are followings:

2.1 Events and Causal Factor Analysis

Events and Causal Factor Analysis is a tool for organizing and analyzing the evidence gathered during an investigation. It is a systematic event analysis tool to aid in collecting, organizing, and depicting event information; validating information from other analytical techniques; writing and illustrating the event investigation report; and briefing on the results of the investigation. The Event and Causal Factor Analysis should be initiated first and updated throughout all root cause investigations. It provides a graphic display of the event on a time line highlighting problems and their causes.

2.2 Change Analysis

Change Analysis is used when the problem is obscure. It is a systematic process that is generally used for a single occurrence and focuses on elements that have changed. As suggested by the name of the tool, change analysis is based on the concept that a change (or difference) can lead to deviations in performance. This presupposes that a suitable basis for comparison exists. What is then required, is to fully specify both the deviated and correct conditions, and then compare the two so that changes or differences can be identified. Any change identified in this process becomes a potential cause of the overall deviation.

Causes identified using change analysis are usually direct causes of a single deviation; change analysis may not yield root causes. However, change analysis may be the only method that can find important, direct causes that are obscure or hidden. This tool of analysis is used in most cases when either the tasks or elements of the task have been completed successfully.

2.3 Barrier Analysis

Barrier Analysis is a systematic process that can be used to identify physical, administrative, and procedural barriers or controls that should have prevented the occurrence. Barrier analysis is based on the concept that hazards represent potentially harmful conditions from which a target (personnel, equipment and environment) must be protected. The purpose of barrier analysis is to identify missing or circumvented barriers. Barrier analysis also shows the barriers that succeeded and prevented the problem from having more serious consequences. The attributes of the barrier analysis tool are: useful to evaluate defense-in-depth and need technically experienced people in the area being analyzed.

2.4 Human Performance Enhancement System

Human Performance Enhancement System (HPES) identifies those factors that influence task performance. The focus of this analysis method is on operability, work environment, and management factors. Man-machine interface studies to improve performance take precedence over disciplinary measures. It is designed directly for investigation of events in nuclear facilities involving human factor related problems and is widely distributed within the nuclear industry.

The HPES method utilizes task analysis, change analysis, barrier analysis, cause and effect analysis and interviewing. Event related information is graphically represented in an event and causal factors chart. The integrated event and causal factors' graphic shows the direct causes, the root causes, the contributing causes, the failed barriers with their interconnections and dependencies.

3 RCA Phases

The objective of investigating and reporting the cause of occurrences is to enable the identification of corrective actions adequate to prevent the recurrence and thereby protect the health and safety of the public, plant personnel and the environment.

The investigation process is used to develop an understanding of the occurrence, its causes and the corrective actions necessary to prevent its recurrence. The line of reasoning in the investigation process outlines what happened step by step. Beginning with the event and identifies the problem (condition, situation, or action that was not called for). Determines what program element was supposed to have prevented this occurrence? Investigate the reasons why this event occurred?

This line of reasoning will explain why the occurrence was not prevented in time and what corrective actions will be most effective. This reasoning should be kept in mind during the entire root cause process. Every RCA process generally comprises of five phases. There may be an overlap between them; efforts should be made to keep them separate and distinct.

3.1 Phase I: Data Collection

The RCA of any event begins with collection of information immediately following the occurrence to ensure there is no loss of vital information and tale-tell signs. Careful preservation of the evidence is very important in the determination of the actual cause of the event. The information collected consists of conditions before, during and after the event; personnel involvement; environmental factors and other relevant information. If the site visit is not possible immediately the area should be quarantined at the earliest and inadvertent entry should be prevented. The investigation team should collect all the relevant information at the earliest, this includes:

- All the documents, procedures, work permits, logs, reports, drawings, etc.
- Interviewing individuals involved in the event at the earliest. The interview should be targeted towards fact finding and not fault finding.
- Data on previous investigations, if any.
- Electronic data on recorders, charts, indicators, etc.
- Photographs of affected area/equipment, etc.
- Samples of gas, effluents for analysis

After data collection all the events are to be mapped on a time line in chronological order to look for any extraneous event or any missing link.

3.2 Phase II: Assessment

Any RCA method may be used that includes following major sThese events are analyzedteps:

- Identify the problem
- Determine the significance of the problem
- Identify the causes or actions immediately preceding and surrounding the problem
- Identify the reasons why the causes in the preceding step existed, working back to root cause.

There are several RCA techniques that can be considered to find the root cause. However, most techniques are effective only in certain situations. Usually a combination of different techniques is required to reach to the root cause. Details of some of the techniques are provided in the preceding section. When performing an RCA that has an element of human error as a part of the cause, it is very important to assume that all individuals do the best job. During the course of the investigation it can be determined if malicious intent was one of the factor. When human errors occur the team members put the errors in proper perspective.

Differentiating between the different causes revealed during an event investigation in order to determine which one is the root cause is a knowledge based skill that requires experience. Once the possible root cause is identified then the next question that comes is whether the event can recur if this cause is permanently corrected. If it can still occur, then the root cause has not been identified.

- Root cause is the most fundamental reason for an event or adverse condition, which if corrected will effectively prevents or minimizes recurrence of the event or condition. A Direct cause is the immediate cause of an event or adverse condition. An Apparent Cause is a cause that can easily be determined by available information without further and deeper investigation.
- Contributing Cause is a causal factor that exacerbated the problem but is not the root cause of the problem.
- Causal Factors are any action or condition either causing an event to occur or increase its severity. A casual factor can be Proximate Root Cause (most probable). There will be cases when the root cause cannot be determined during the investigation due to lack of sufficient data, inability to identify the exact failure, or a delay in revealing the failure due to outages or extended failure analyses. In these cases, the Proximate Root Cause should be determined. The proximate root cause is the best root cause that can be determined based on all of the information available.

3.3 Phase-III: Corrective Actions

Implementing effective corrective actions for each cause reduces the probability that a problem will recur and improves reliability and safety. Corrective Actions are taken to address the root causes of issues (Equipment, Organizational, human performance, etc.). If the corrective actions addressed by the analysis could not be taken immediately, then Interim Actions are called for. Interim Actions are important for mitigating or preventing the effects of the causes until Corrective Actions to Prevent Recurrence can be fully implemented. Interim Actions are sometimes implemented immediately upon discovery of the event, or they can be initiated at any time throughout performance of the root cause investigations.

3.4 Phase-IV: Reporting

Typically the complete analysis report should be submitted within a month depending upon the severity of the event and further detailed analysis. The report should include:

- The initiating event report
- Investigation terms of reference

- Equipment failure worksheet
- Failure analysis report
- Root cause investigation report
- The corrective actions

The report should be presented in the safety committees by the RCA team.

3.5 Phase-V: Follow-up

Follow-up includes determining if corrective actions have been effective in resolving problems. An effectiveness review is essential to ensure that corrective actions have been implemented and are preventing recurrence.

3.6 RCA Methodology—BARC

At Research Reactors in BARC, each occurrence termed as an 'Event' is reported by plant Operations and discussed in plant safety committee. Depending upon the importance of the event to the safety of the plant, it is categorized as either a 'Low Level Event' or 'Significant Event'. The Low Level Events fall under Level-1 RCA category and more severe Significant Events are categorized under Level-2 RCA category. The general approach adopted for analysis of both types of events is described below:

3.7 Level 1 RCA (Low Level Events)

As described earlier these types of events pose no safety hazard and keep occurring in large numbers. Different types of events occurring in a plant and their frequency is represented in Fig. 1 [1]. These events are analyzed in an intense brainstorming session among the team members of the RCA committee and a domain expert if

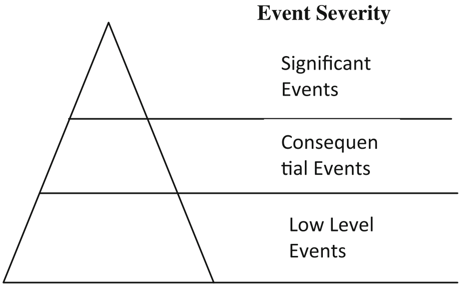


Fig. 1 Frequency of occurrence of different types of events

required; the solution or conclusion is arrived after one discussion session. If required, a second brainstorming session is held after gathering enough data about the failure or event and the corrective actions are chalked out. The analysis is performed by following a single RCA method or at the most two different methods. The RCA report in such cases is prepared within a fortnight and submitted to Plant Safety Committee. If repetition of some of the events or large number of similar events is observed then a detailed analysis is performed and the analysis may take a Level-2 route. This methodology is also employed to analyse 'precursor' events. The precursor events are the latent weaknesses identified from plant SSCs before they lead to a serious event. The PSA methods make it possible to quantify the likelihood that a precursor event will turn into a serious event or an accident.

3.8 Level-2 RCA (Significant Events)

The significant events are safety related events and the plant submit a prompt notification to the regulator. These events are categorized as 'Anomaly or Deviation' in the IAEA International Nuclear Event Scale (IAEA-INES) and rank below 'Zero'. In such cases every attempt is made to collect all the relevant information about these events at the earliest. The analysis of these types of events requires several discussions among the members and if need arises an expert from the relevant field is invited. Investigations of some of the event requires further

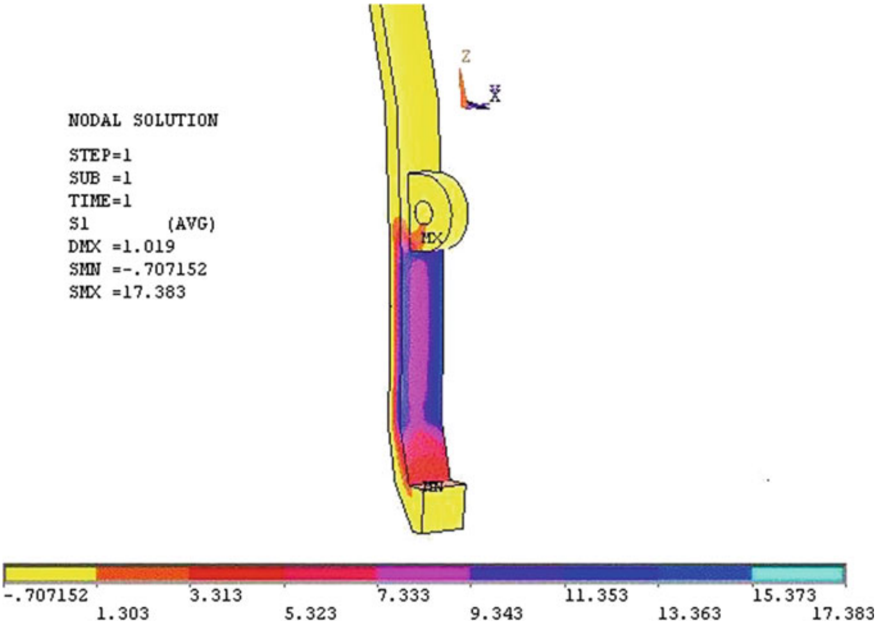


Fig. 2 FEM model of a fuel assembly grappler jaw

detailed analysis in the area of Reactor Physics, Metallurgy and Metallographic investigations, Material Characterization, Component Stress Analysis using FEM modeling, Chemical Analysis, etc. For performing these types of detailed investigations as well as accelerated life tests of components a 'Life Cycle Reliability Engineering Laboratory' at Dhruva, BARC has been established with the necessary expertise. The lab houses a Thermal-Humidity chamber for life assessment studies, Thermal Imaging Camera, a Scanning Electron Microscope for metallographic studies and material characterization, etc. An attempt is made to issue the detailed investigation report within a month's time or a preliminary report is issued if results of some of the analysis are pending. An FEM model of a fuel assembly grappler jaw developed during performing RCA of failure of the grappler jaw is shown in Fig. 2 [4].

Table 1 Barrier analysis worksheet

Undesirable event	Existing barriers	Barrier failure (Yes/No)	How barrier failed?	Why barrier failed
Fire in plant during plasma cutting	Administrative Barrier: Welding/gas cutting permit procedure	Yes	The precautions given in the gas cutting permit were not followed	Unaware about the presence of flammable material in the area. (Deviation from permit procedure)
	Physical barrier: Asbestos cloth used as protection from three sides	Yes	The asbestos cloth used as a physical barrier had small holes & minute openings	Physical barrier inadequate
			The physical barrier only covered the three sides and not covered the top. Reflected spatters can escape from top	
Reasons for delay in detection of fire				
Delay in detection of fire	Environmental barrier:	Yes	Smoke generated by plasma cutting accumulated in the area	The poor ventilation in the area led to accumulation of smoke and the poor illumination led to poor visibility. So the fire could not be promptly located
	Visibility in the area		The area operator after getting information about fire alarm visited the area, noticed smoke in the room & noted people working in the room	
	Fire & beetle alarm annunciation			
Delay in informing control room	Administrative barrier: Fire emergency procedure	Yes	The fire emergency procedure was not followed	It appears that the staff available at site got panic-stricken & started fire fighting operation

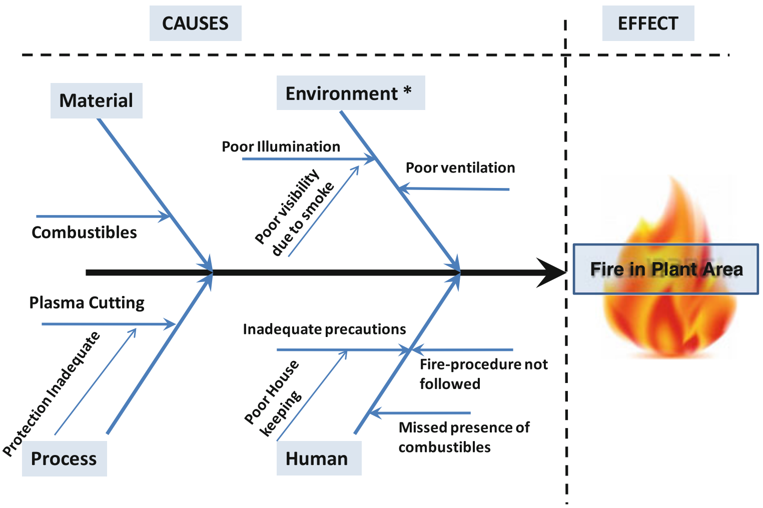


Fig. 3 Fish bone diagram for fire event (*Environment factor resulted into delayed detection of the fire event)

4 Case Study

A brief case study of a fire event that had taken place in the plant is presented here. During routine shutdown of the reactor a modification job was taken in the plant. The job involved plasma cutting of SS sheets. During this process the spatters generated initiated a fire as some combustible material was present in the area. The event was analysed by the committee and a Barrier Analysis work sheet as shown in Table 1 along with a Fish-bone diagram shown in Fig. 3 are presented here.

5 Conclusion

In a Nuclear Plant the occurrence of different types of events is unavoidable however, they can be controlled. If efforts are put to record and analyze each event, the lessons learnt from past instances will lead to control and reduce further occurrences. This will promote the safety culture at plant level.

The RCA approach employed for the components failure study reveals that the cause of the failure can lay in its faulty design or inadequate maintenance practices. Reliability based safety assessment tools such as Probabilistic Safety Assessment (PSA) has been very helpful in assessing the safety level of plant. The minimum cut sets provided by the PSA can identify the shortest path by which a component failure can propagate, degrade the system and deteriorate the safety. Rectification of the root causes obtained by the insight of the detailed RCA is helpful in increasing the reliability of the components in general and in particular the components belonging to minimum cut sets list and hence enhancing system reliability. RCA will help in screening raw failure data before deriving useful reliability data for PSA.

References

1. Trending of low level events and near misses to enhance safety performance in nuclear power plants—IAEA TECDOC
2. Root Cause Analysis Following an Event at a Nuclear Installation: Reference Manual-IAEA TECDOC 1756
3. DOE-NE-STD-1004-92, Guideline, Root Cause Analysis Guidance Document 1992), US Department of Energy, Office of Nuclear Energy, Washington DC, USA
4. Varde PV RCA of precursor event: FMB grappler jaws failure—a BARC internal report

Maintenance of Large Engineering Systems

Anil Rana, Ajit Kumar Verma and Ajit Srividya

Abstract Maintenance has been defined as the combination of all technical and administrative actions, including supervision actions, intended to retain an item in, or restore it to, a state in which it can perform its required function. It is a set of organised activities that are carried out in order to keep an item in its best operational condition with optimal utilisation of resources. In a survey conducted by the author regarding use of resources in two different naval commands, it was estimated that over 50–60 % of total operation cost went into maintenance of the ships and its machinery. Though the operation cost could be directly linked with the achieved operational objectives it was extremely difficult to justify the maintenance cost with its accrued benefits.

Keywords Maintenance · Large engineering systems · Operational cost · Maintenance optimization

1 Background

Over the years, maintenance has been given the due place it deserves in many large industrial setups and that includes the shipping industry. While in the past, maintenance was only seen as an additional cost factor, [1] both in terms of loss of opportunity and utilization of resources, it is only post the 1980s that it is being considered as the one of the most invaluable part of the organization which can have a direct impact not only on cost savings but also on improvement in safety, quality and reliability. This realization of its importance initiated a lot of work into development of maintenance optimization models, all of which aimed at a common goal of improving the system reliability or availability in a cost effective manner.

A. Rana (✉)
Fiji National University, Suva, Fiji
e-mail: anil.rana@fnu.ac.fj

A.K. Verma · A. Srividya
University College, Haugesund, Norway

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_43

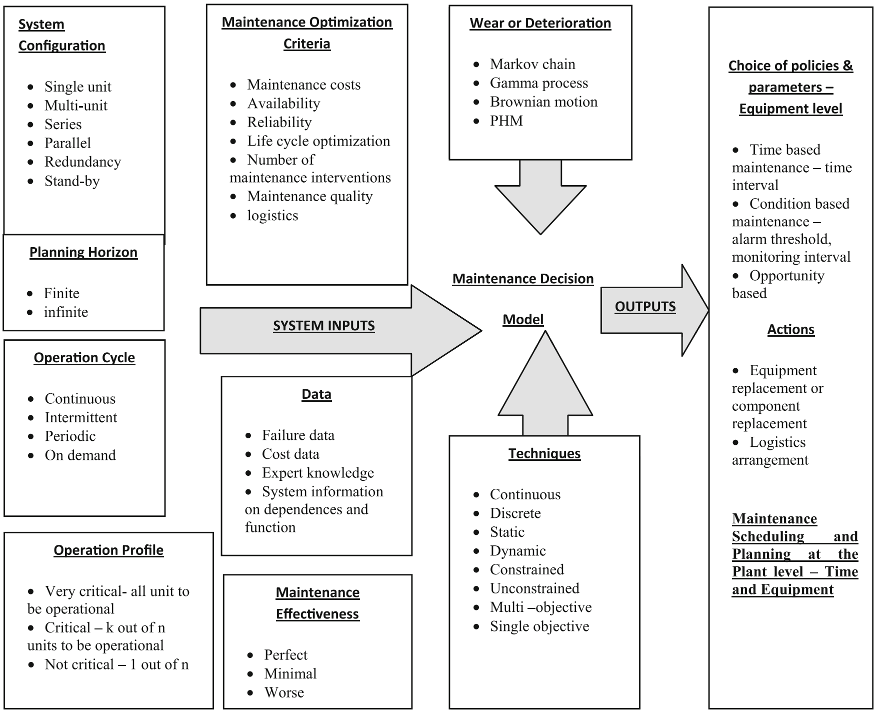


Fig. 1 An overview of a maintenance optimization model

This movement got a fillip with the improvement in technologies pertaining to information processing and analysis, machine health and condition monitoring, data and inventory management and as a result, more and more sophisticated maintenance policies and strategies have been developed which have maximized the probability of realizing the objectives of maintenance. Though a general optimization model which fits all the needs of all the maintenance aspects of organization is difficult to come by, a general understanding of its inputs and requirements based on the envisaged objectives, its effectiveness and its configuration has already emerged. An overview of the different aspects of a maintenance optimization model is shown in Fig. 1 Horenbeek [2].

It is surprising therefore that in spite of the realization of the importance of maintenance as a function of any organization and emergence of an understanding of the general maintenance models, there still exists a large gap between the maintenance that is being practiced onboard ships and the one that is being preached in theory.

2 Gap in Theory and Practice

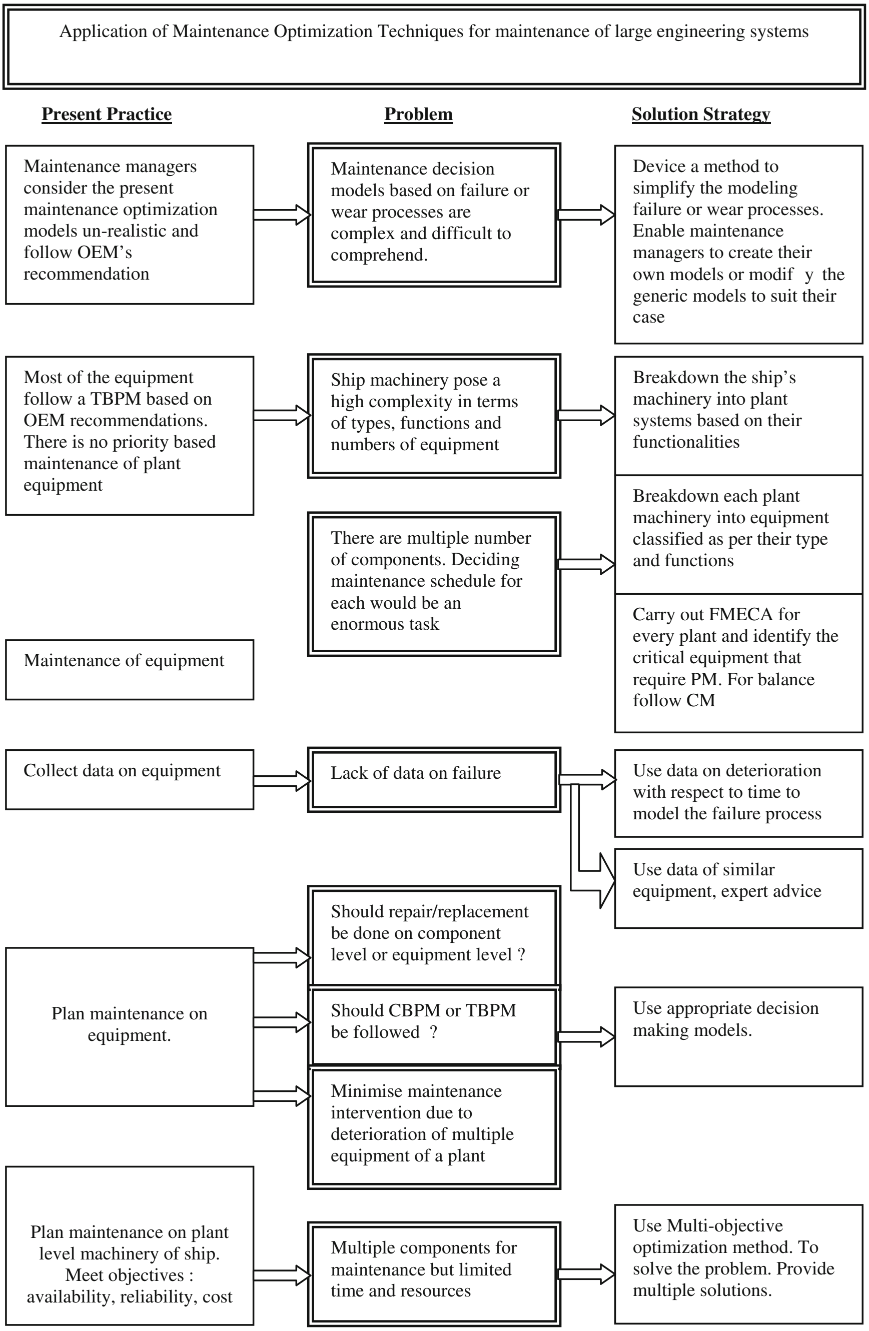
Authors such as Dekker [3, 4], Scarf [5] and Garg et al. [6] have collectively surveyed more than 300 papers on maintenance models. It has been brought out by them that most of the research work on the subject has been carried out at an individual component or equipment level. Horenbeek et al. [2] bring out that case studies are often used only to demonstrate the applicability of a developed model, rather than finding an optimal solution to a specific problem of interest to a practitioner. Nicolai and Dekker [7] conclude that case studies actually faced by the maintainers of the equipment are not well represented in literature. Dekker [3] attributes this to lack of motivation on the part of the practitioners and also to the complexity of the underlying optimization models. A case in point is the phenomenon of wear or deterioration. Though this phenomenon is commonly observed in mechanical systems onboard, there is still a reluctance to mathematical model it in terms of a gamma wear simply for the reason that it is analytically difficult to handle it.

Another limitation perceived in literature is that most of the models focus on only one optimization criterion, making multi-objective optimization models an unexplored area of maintenance optimization. Although single objective optimization is attractive from the modeling point of view, this approach does not capture all important aspects of a real life situation. Surveys carried out by eminent researchers have brought out the fact that the other reason for this wide gap between theory and practice is that most of the maintainers find it difficult to understand the complex mathematical models most of which are written for mathematical purposes only with little regard to its applicability.

The need of the hour therefore is to take actions on two fronts: firstly, to generalize the maintenance decision problems into broader groups based on their common characteristics and create mathematical models that are realistic and can be systematically applied to ‘on field’ situations with little changes. Secondly and more importantly, to develop a simpler tool for modeling the intricate failure processes of equipment so that it can be used by the maintenance personnel on field. The tool can then help the maintenance personnel prepare their own maintenance decision models (based on failure processes) for analysis and in addition it would make them capable of even altering the generalized models available in literature to suit their own specific requirements.

3 Problem Formulation and Solution Strategy

The problem formulation and solution strategy are graphically demonstrated in Figs. 2 and 3 along with a step by step solution process. The strategy and the solution process are self-explanatory. The main concerns handled by the author, so far, in terms of systematic treatment of optimization of maintenance actions are



Application of Maintenance Optimization Techniques for maintenance of large engineering systems

Present Practice

Problem

Solution Strategy

Maintenance managers consider the present maintenance optimization models un-realistic and follow OEM's recommendation

Maintenance decision models based on failure or wear processes are complex and difficult to comprehend.

Device a method to simplify the modeling failure or wear processes. Enable maintenance managers to create their own models or modif y the generic models to suit their case

Most of the equipment follow a TBPM based on OEM recommendations. There is no priority based maintenance of plant equipment

Ship machinery pose a high complexity in terms of types, functions and numbers of equipment

Breakdown the ship's machinery into plant systems based on their functionalities

There are multiple number of components. Deciding maintenance schedule for each would be an enormous task

Breakdown each plant machinery into equipment classified as per their type and functions

Maintenance of equipment

Carry out FMECA for every plant and identify the critical equipment that require PM. For balance follow CM

Collect data on equipment

Lack of data on failure

Use data on deterioration with respect to time to model the failure process

Use data of similar equipment, expert advice

Plan maintenance on equipment.

Should repair/replacement be done on component level or equipment level ?

Should CBPM or TBPM be followed ?

Minimise maintenance intervention due to deterioration of multiple equipment of a plant

Use appropriate decision making models.

Plan maintenance on plant level machinery of ship. Meet objectives : availability, reliability, cost

Multiple components for maintenance but limited time and resources

Use Multi-objective optimization method. To solve the problem. Provide multiple solutions.

Fig. 2 Problem formulation and solution strategy

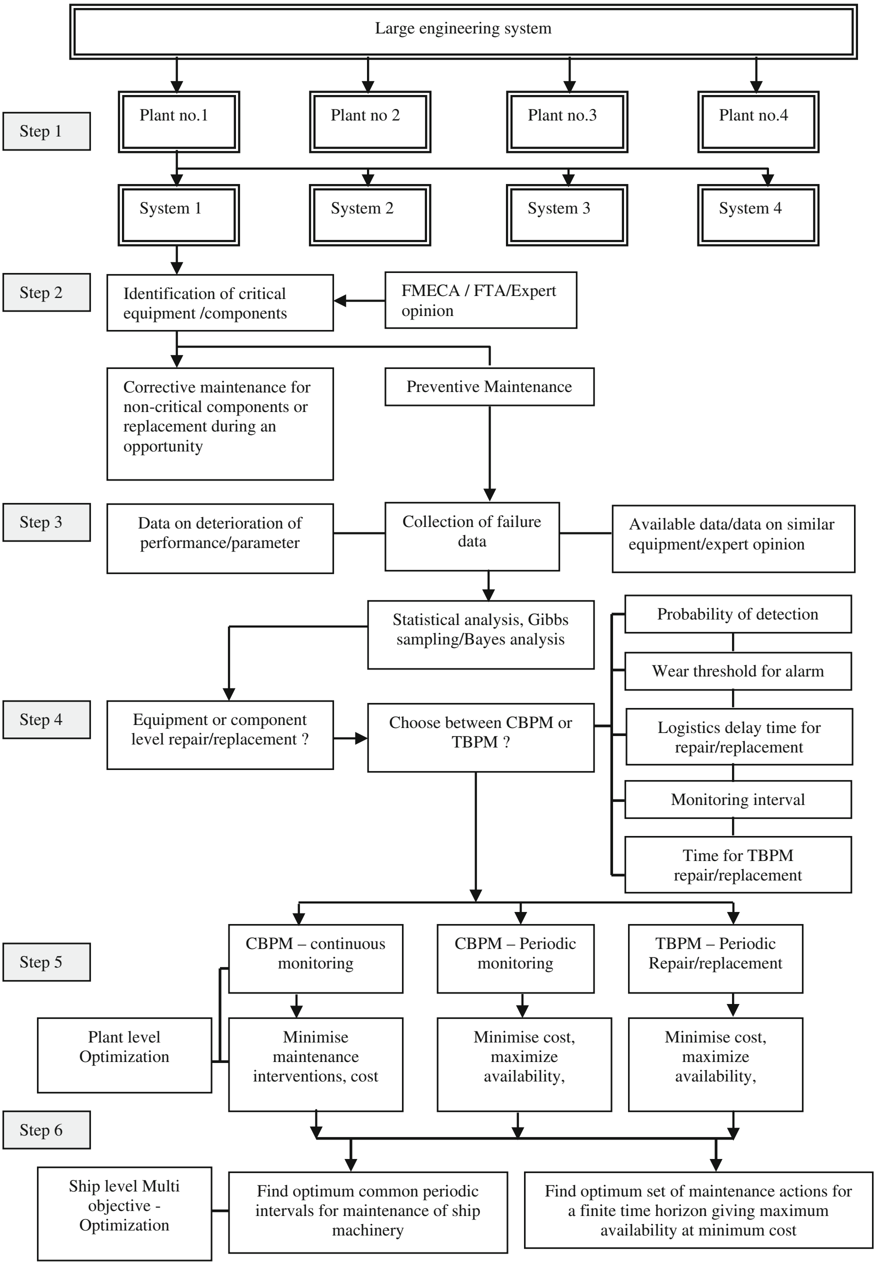

Large engineering system
Step 1
Plant no.1
Plant no 2
Plant no.3
Plant no.4
System 1
System 2
System 3
System 4
Step 2
Identification of critical equipment /components
FMECA / FTA/Expert opinion
Corrective maintenance for non-critical components or replacement during an opportunity
Preventive Maintenance
Step 3
Data on deterioration of performance/parameter
Collection of failure data
Available data/data on similar equipment/expert opinion
Statistical analysis, Gibbs sampling/Bayes analysis
Probability of detection
Wear threshold for alarm
Step 4
Equipment or component level repair/replacement ?
Choose between CBPM or TBPM ?
Logistics delay time for repair/replacement
Monitoring interval
Time for TBPM repair/replacement
Step 5
CBPM – continuous monitoring
CBPM – Periodic monitoring
TBPM – Periodic Repair/replacement
Plant level Optimization
Minimise maintenance interventions, cost
Minimise cost, maximize availability,
Minimise cost, maximize availability,
Step 6
Ship level Multi objective - Optimization
Find optimum common periodic intervals for maintenance of ship machinery
Find optimum set of maintenance actions for a finite time horizon giving maximum availability at minimum cost

Fig. 3 Step by step solution process

listed below. It may be noted that the constraints of availability of onboard space, spares and expertise adds special dimension to the problem that need to be resolved by timing the maintenance actions along with the scheduled visits of the ships to harbours.

- A framework for taking maintenance decision at different hierarchical level of plant machinery was identified and published [8]. FMECA (failure mode effect and criticality analysis) was used to prioritize the equipment that could be treated optimally.
- At equipment level, a framework to decide between a CBPM (condition based predictive maintenance) or a TBPM (time based preventive maintenance) policy was published [9].
- A framework was presented to choose the level of repair/replacement (whether equipment level or component level) based on cost of components, cost of equipment, cost of repair and failure distributions of the components [10].
- Maintenance action plan for a wide variety of equipment for a particular class of ship was published based on an evolutionary algorithm. It has been demonstrated that grouping of some select equipment for maintenance during a short maintenance period, scheduled in between the refit period of a ship is beneficial [11]
- A framework to model the maintenance decision requirement for a large plant with multiple equipment and different times in use, as a MOOP (multi-objective optimization problem), based on cost and availability and with constraint of time and resources was presented [11].
- Wear was successfully modeled as a gamma process. The advantage of this process is that it considers the wear or deterioration of equipment/component in its various stages as against a Weibull process that simply labels equipment in an 'on or off' state. The gamma process is therefore more realistic and also amenable to simplification using a stochastic petrinet methods explained later in a paper. The greatest benefit of the process lies in use of data observed by a ship machinery operator for ascertaining the gamma process parameters. For example, consider the data on a steam turbine of a ship given in Table 1.

The observed data were analysed using various techniques including Gibbs sampling (using WINSBUG program) to arrive at the gamma process parameters. If the shape parameter is a function of time $\lambda(t) = \alpha.t^{\zeta}$ and the scale parameter is β then pdf (probability density function) is $f_{X(t)}(x) = \frac{\beta^{\alpha.t^{\zeta}}}{\Gamma(\alpha.t^{\zeta})} x^{\alpha.t^{\zeta}-1}.e^{-\beta x} = Ga(x|\lambda(t), \beta)$,

The parameters obtained can then be used to arrive at mean time to failure dependent on states of wear or deterioration [12, 13]. The Gibbs sampling analysis of the recorded data by ship operator and its simulated gamma process are shown in Figs. 4 and 5 below

Table 1 Wear data of turbine components

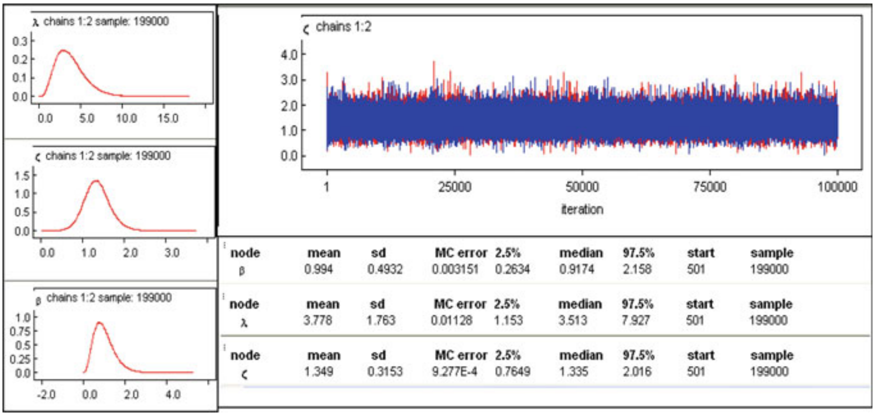
Bearings		Labyrinths		Diaphragm		Cam/Nozzle assy	
Time	Wear	Time	Wear	Time	Wear	Time	Wear
0.45	1.77	0.1	0.00001*	0.45	1.11	0.5	2.75
1	2.44	0.5	6.889	1	2.22	0.9	7.75
1.52	3.422	0.7	7.33	1.52	4.44	1.1	8.75
1.8	3.98	0.75	7.44**	1.8	4.45	1.4	9
2.85	5.16	1.2	8.6	2.85	5.56	1.6	9.125
3.4	6	1.4	8.8667	3.4	6.67		
3.65	5.97***			–	–		
3.9	6.9			3.9	7.8		
4.2	7.3			4.2	8.89		
4.5	8.07			4.5	10		

All data have been normalised to a scale between 1 and 10 non-dimensional parameter

* The wear could not be measured, but an arbitrary small value has been mentioned to facilitate calculation of the distribution parameters using MLE (max likelihood estimate) method

** Wear data are each for different sets of labyrinths (except the first two serials). The 3rd serial was actually 7.111

*** Data rejected as wear could not have reduced



```
node    mean    sd      MC error  2.5%    median  97.5%   start   sample
 β      0.994   0.4932  0.003151  0.2634  0.9174  2.158   501     199000

node    mean    sd      MC error  2.5%    median  97.5%   start   sample
 λ      3.778   1.763   0.01128   1.153   3.513   7.927   501     199000

node    mean    sd      MC error  2.5%    median  97.5%   start   sample
 ς      1.349   0.3153  9.277E-4  0.7649  1.335   2.016   501     199000
```

Fig. 4 Estimates of Gamma distribution parameters through Gibbs Sampling

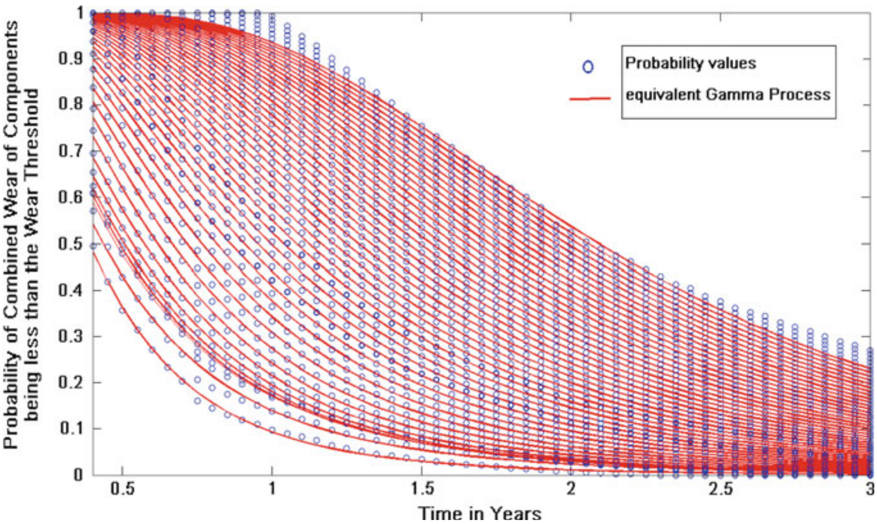


Fig. 5 Plot of an Equivalent gamma process (1-CDF) fitted on to a probability curve

4 Simplification of Maintenance Modeling

As brought out earlier, a need for simplification of maintenance modeling has been felt by various industries including the shipping industry that would allow the maintenance managers in the field to effectively apply these models in their specific areas of applications. An already existing tool called the Petrinets was researched by the author. Petrinets are known to be a very useful graphical, mathematical modeling and analysis tool applicable to a variety of applications such as queue performance evaluations, communication protocols, in-process monitoring systems, real time fault tolerant and safety critical systems etc. [14–24]. However, the limitation of use of memory-less exponential random variables render this tool difficult to be used for modeling age based wear and failure processes.

The above problem was overcome by adopting the method of stages Cox et al. [25] to convert the non-exponential processes into a mixed Erlang process. Such a conversion facilitated the age memory of the processes and made it amenable to be solved through computer based simulation programs without the need of introduction of an ‘age variable’ to keep track of the elapsed time. The method has been successfully used to demonstrate the failure processes of systems, such as stern tube bearing assembly of a ship, safety device failures, two pump based auxiliary system etc. [26, 27] through comparison between the simulation studies and mathematical model based results. The highlights of this technique are:

- The method simplifies the modeling of failure processes of the mechanical systems (with age memory) and thereby makes it amenable for use by the maintenance engineers on field.

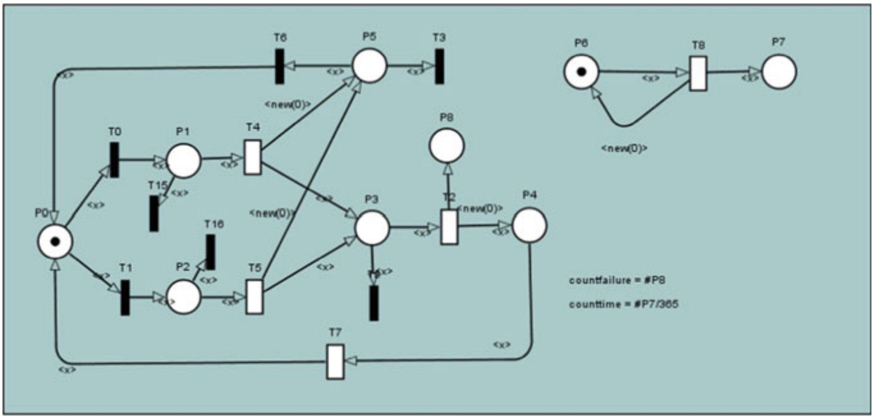


Fig. 6 Petrinet model for Gamma wear process in Timenet

- The method is especially beneficial to handle wear or deterioration based processes.
- The method also helps in carrying out simulation studies on multi-component system where the petrinet models generated using the above method act as building blocks.

A stochastic petrinet model in "Timenet" that can simulate the gamma wear process is shown in Fig. 6.

5 Conclusion and Future Work

The broad based maintenance models characterized by similar decision problems and similar hierarchical levels of machinery set ups, as outlined in the solution strategy, were applied to 2 classes of ships in a particular naval command. Though the benefits were not immediately quantifiable for documentation, but the trend was positive. There were some direct savings in maintenance costs and logistics support. The impact on reliability and availability of the ships was only possible to be studied over a period of 3 years (at the minimum).

As regards the simplification of maintenance models, stochastic Petrinets have shown to be a potent tool after suitable modification as brought out by the author in various publications. However, as the size of the components increase in number it no longer remains feasible to handle it analytically. There is therefore a need to prepare a software exclusively to handle maintenance related problems which, as brought out earlier, are large in variety and specific to the configuration and composition of the equipment.

References

1. Shenoy D, Bhadury B (1998) Maintenance resources management: adapting MRP. Taylor and Francis, New York, pp 5–6
2. Horenbeek VA, Pintelon L, Muchiri P (2010) Maintenance optimization models and criteria. In: Proceedings of the 1st international congress on e-maintenance, pp 5–13, Lulea Sweden
3. Dekker R (1996) Applications of maintenance optimization models: a review and analysis. Reliab Eng Syst Saf 51(3):229–240
4. Dekker R (1995) On the use of operations research models for maintenance decision making. Microelectron Reliab 35(9–10):1321–1331
5. Scarf PA (1997) On the application of mathematical models in maintenance. Eur J Oper Res 99:493–506
6. Garg DS (2006) Maintenance management: literature review and directions. J Qual Maint Eng 12(3):205–238
7. Nicolai R, Dekker R (2007) Optimal maintenance of multi-component systems: a review, complex system maintenance handbook: blending theory with practice. Springer, Berlin
8. Verma AK, Srividya A, Rana A (2012) Optimization of maintenance scheduling of ship borne machinery for improved reliability and reduced cost. Int J Reliab Qual Saf Eng 19(03)
9. Verma AK, Srividya A, Rana A (2013) Search for optimal preventive maintenance policy of equipment under an uncertainty of detection of its condition. SRESA's Int J Life Cycle Reliab Saf Eng 2(2)
10. Rana A (2011) Use of rotables—Repair at component or equipment level. J Marine Eng Indian Navy 64:12–16
11. Rana A (2012) PhD Thesis on multi-objective optimization of maintenance of ship borne machinery and an approach to its analysis using stochastic Petrinets. Indian Institute of Technology, Mumbai, India
12. Verma AK, Srividya A, Rana A (2011) Approximation of MTTF calculation of a non-stationary gamma wear process. Int J Syst Assur Eng Manag 2:282–285
13. Verma AK, Srividya A, Rana A (2011) Optimal time scheduling for carrying out minor maintenance on a steam turbine. Int J Syst Assur Eng Manag :1–12
14. Ajmone Marsan M, Conte G, Balbo G (1984) A class of generalized petrinets for the performance evaluation of multiprocessor systems. ACM Trans Comput Syst 2:93–122
15. Ammar HH, Huang YF, Liu RW (1987) Hierarchical models for systems reliability, maintainability and availability. IEEE Trans Circuit Syst 34:629–638
16. Holiday MA and Mary K Venon, Dec 1987, A generalised time Petri net model for performance analysis. IEEE Trans Softw Eng SE-13:1297–1310
17. Hass PJ, Shedler GS (1989) Stochastic petrinet representation of discrete event simulation. IEEE Trans Softw Eng 15:381–393
18. Molloy MK (1982) Performance analysis using stochastic Petrinets. IEEE Trans Comput C-31:913–917
19. Ramamoorthy CV, GS Ho (1980) Performance evaluation of asynchronous concurrent systems using Petrinets. IEEE Trans Softw Eng SE-6:440–449
20. Billington J, Wheeler GR, Wilbur-Ham MC (1988) PROTEAN: a high level petrinet tool for the specification and verification of communication protocols. IEEE Trans Softw Eng 14:301–331
21. Chaillet A, Combacau M, Courvoisier M (1993) Specification of FMS real time control based on petrinets with objects and process failure monitoring. Proc IECON-93, Hawaii, pp 144–149
22. Leveson NG, Stolzy JL (1987) Safety analysis using Petrinets. IEEE Trans Softw Eng 13:386–397
23. Srinivasan VS, Jafari MA (1993) Fault detection/monitoring using timed petri nets. IEEE Trans Syst Man Cybern 23:1155–1162
24. Petri CA (1962) Kommunikation mit Automaten. Ph.D Thesis, University of Bonn, Federal Republic of Germany

25. Cox DR, Muller HD (1970) The theory of stochastic processes. William Clowes and Sons, London
26. Verma AK, Srividya A, Rana A (2014) Use of stochastic Petrinets in modeling of safety device inspection interval problem. Int J Syst Assur Eng Manag. doi:10.1007/s13198-014-0264-z
27. Verma AK, Srividya A, Rana A (2011) Use of Petrinets for solution of a Stern gland optimal inspection interval problem. Int J Syst Assur Eng Manag 2:183–192

Maintenance Risk Based Inspection Optimization Model in Multi-Component Repaiable System with Economic Failure Interaction

Esmaeil Rezaei and Din Mohammad Imani

Abstract Inspection is one of the important activities to detect and fix failures in repairable system. Optimization of inspection intervals decreases expected costs of maintenance system, reduces inspection costs and increases the performance of operating system. This paper proposes a new model to reliability and cost evaluation and figure out the optimal periodic inspection interval on a finite time horizon. In addition, the grouping maintenance preferred to individual maintenance as economic failure dependency. The costs are include inspection, repair, and downtime. The downtime cost is proportional to the elapsed time from failure time to its detection at next inspection time. On a finite time horizon, the objective of current study is to figure out the optimal inspection interval for the soft failure component to minimize the expected total cost. In addition, the expert judgment is used to considering risk in inspection. Therefore, a sample problem is solved and numerical results are presented. Result indicates, the risk reduces inspection interval time.

Keywords Periodic inspection interval · Risk based inspection (RBI) · Repairable system · Soft and hard failures · Failure interaction

1 Introduction

The cost of maintenance is one of the major performance indexes in manufacturing and operation. The aim of maintenance is to maximizing reliability and minimizing cost [1, 2]. The optimization cost can be considered as returning potential lost profit or budget injection in maintenance [3, 4]. Also repairing and inspections increases

E. Rezaei (✉)
Mazandaran University of Science and Technology, Behshahr, Iran
e-mail: esmaeil.rezaei@b-iust.ac.ir

D.M. Imani
Iran University of Science and Technology, Tehran, Iran
e-mail: imanim@iust.ac.ir

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_44

maintenance costs. In contrast, increasing inspection and maintainability reduces the downtime penalty cost. The inspection is one of the important activities to detect and fix failures in repairable system. Optimization of inspection intervals has critical role in decreasing total costs of maintenance system, reduces inspection costs and increases the performance of operating system.

Risk Based Inspection (RBI) is one of the most important aspect of maintenance [3–6]. Risk has critical role in maintenance and inspection interval. For example, failure of transformer is few. But its inspection intervals time is short due to risk. So, the risk plays important role in inspection interval optimization problem.

In inspections activity, failures of system are investigated. The failure of components can be soft or hard. The hard failure causes the system stop, while the soft failure does not, but it increases the system operating costs too.

A great deal of periodic inspection research for hidden failure addresses the cost issue. In these works, the optimum inspection interval and maintenance policy are obtained by minimizing the expected cost in a given time period [7].

Barlow et al. [8] have created a basic model, for determining the optimal inspection interval to minimize the expected cost. Accordingly, in their model, only two costs are included: a fixed inspection cost for each inspection; and a loss (downtime) cost per unit time resulting from the elapsed time between failure occur and the failure detection (next inspection). While the Barlow et al. model only considered inspection cost and loss cost, other research takes the repair/replacement cost into account [3, 9, 10]. Taghipour and Banjevic [10] investigate the optimal inspection interval for a multi-unit repairable system to minimize total expected cost over a finite and infinite time horizon. Ahmadi and Kumar [11] develop a cost rate function model to determine the optimum inspection interval time and frequency of inspection and restoration of an aircraft's repairable components.

In a multi-component system, the failure of one components may interact with another. These interactions create dependency among the components and categorized as follows [3, 12]:

Economic dependence, occurs when the cost of maintenance and replacement creates dependency among the components. In other words, the grouping maintenance may cost less than the maintenance them individually [13]. This type of dependency is the focus of the present study.

Structural dependence, occurs when the maintenance and replacement of some components require replacement or disassembly of some other parts or components [12].

Stochastic/Probabilistic dependence, which happens when the state of a component, such as its workload, affects the life-time distribution of the other components. For instance, the failure of one component increases the failure/hazard rate of other components [12].

In 1986, Thomas [12] put together a survey reviewing the models which were previously proposed for complex systems along with their maintenance and replacement policies. In 1991, Cho and Parlar [14] reviewed the maintenance of various multi-component models. In 2011, Sarkar et al. [15] reviewed the literature and collected different maintenance policies for complex systems. This research provides a good overview for both single and multi-component systems during the past 50 years. According to these reviews [12, 14, 15], while there are several publications on multi-component systems with economic dependence, the studies on complex systems with stochastic dependence are sparse. Most of these studies only consider two-component systems, because in practice it is difficult and sometimes impossible to evaluate the actual effect of the failure of multiple components on each other [16]. Murthy and Nguyen [17, 18] studied the maintenance of systems considering stochastic dependence. They formulated the failure interactions between components in a two and multi- component systems and developed expressions for the expected operation costs for both finite and infinite life-times. Scarf and Deara [19] developed a model considering both economic and stochastic dependences between components in a two-component system. The policies in their model were age-dependent. They further extended their model to block replacement policies for a two component system [20]. Zequeira and B'erenguer [21] analyzed the maintenance costs for a two-component standby parallel system taking into account the stochastic dependence. In their study, they considered periodic inspections and preventive maintenance. Taghipour et al. [22] proposed a model to find the optimal periodic inspection interval on a finite time horizon for a complex repairable system. They considered costs include inspection, repair, and downtime penalty cost. Inspection interval with failure interaction for two and multi components have been studied by [23, 24]. They considered a two-component system. In their studies, the capacitor bank (first component) and the transformer (second component) for a distribution substation in an electric power distribution system were considered. Recently, Rezaei and Imani [3] proposed new risk base inspection optimization model to by considering fuzzy failure interaction. Their considered probabilistic dependency. They applied Simpson rules and Bayesian theory to modeling and solving. Their model required less calculation in compare to [22].

Recently, Rezaei and Imani [1] proposed new risk basedinspection optimization model by considering fuzzy failure interaction. They assumed, the system can be worse along failure occur to failuredetection (next inspection) and followed the minimal repair policy.They also, studied optimization inspection interval under perfect repair policy [2].

Proposed Inspection Optimization Model

The problem definitions and assumptions are maintained in introduction. In this section, the proposed model presents. At the first, the parameters and variables definition presents as Table 1.

The components failures generally fallow as weibull distribution. See Eq. (1) for the hazard rate of weibull distribution:

$$\lambda(x) = \frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1} \tag{1}$$

The Cumulative distribution function is given by Eq. (2) and simplification by equations, Eq. (3).

$$F_i(x) = 1 - e^{-\int\limits_{t}^{t+x} \lambda_i(x)dx} \quad 0 \leq x \leq \tau \tag{2}$$

$$F_i(x) = 1 - e^{-\int\limits_{t}^{t+x} \frac{\beta_i}{\theta_i}\left(\frac{x}{\theta_i}\right)^{\beta_i - 1} dx} \overset{t=0}{=} 1 - e^{-\left(\frac{x}{\theta_i}\right)^{\beta_i}} \tag{3}$$

The $p^k_{\tau,i}(t)$ is probability that the component i doesn't fail in kth inspection interval of the cycle T with τ inspection interval that defined as reliability function. To obtain reliability, $p^k_{\tau,i}(t)$, the Bayesian theory is used. Due to different times of inspection intervals, the probability of $p^k_{\tau,i}(t)$ is depends on $p^{k-1}_{\tau,i}(t)$. The Bayesian approach to obtain $p^k_{\tau,i}(t)$ is given by

$$\begin{aligned} p^k_{\tau,i}(t) &= p^k_{\tau,i}(t|safety\, in\, p^{k-1}_{\tau,i}(t))p^{k-1}_{\tau,i}(t) + \\ & p^k_{\tau,i}(t|unsafety\, in\, p^{k-1}_{\tau,i}(t))(1 - p^{k-1}_{\tau,i}(t)) \quad ,k = 1, \ldots, T/\tau \end{aligned} \tag{4}$$

Table 1 Parameters and variables definition

$\lambda_i(x)$	The failure rate of the soft component i at time x	C_i^s	The cost of each inspection of the component *i*
		C_i^d	The cost of each perfect repair of the component *i*
$((k-1)\tau, k\tau]$	kth inspection interval in the cycle T, *k* = 1, 2, …, *n*	C_i^p	The downtime penalty cost associated with the component *i* from failure occur to its detection at the inspection time
$\tilde{\psi}$	fuzzy risk parameter		
$E\left[C^\tau\right]$	The expected total cost of the all components in the inspection interval τ	*n*	The number of inspections to be performed on the soft component during the cycle T
$E\left[C_i^{(k-1)\tau,k\tau}\right]$	The expected total cost of the component *i* in *k*th inspection interval of the cycle T, i.e. From a scheduled inspection at $k\tau$ over time period $((k-1)\tau, k\tau]$.	T	The planning horizon length (e.g. 1 year) which is known and fixed
τ	The time between two consecutive inspections, $\tau = T/n$	t	The initial age of the soft component at the beginning of the cycle T
$p^k_{\tau,i}(t)$	The probability that the component *i* doesn't fail in *k*th inspection interval of the cycle T with τ inspection interval (reliability function), provided that we know that its age at the beginning of the cycle T is equal to t .	m	Number of components

From Eqs. (4) and (2)

$$p^{k}_{\tau,i}(t) = p^{k-1}_{\tau,i}(t)\left[1 - F_i(x)|^{k\tau}_{(k-1)\tau}\right] + (1 - p^{k-1}_{\tau,i}(t))\left[(1 - F_i(x)|^{\tau}_{0})\right] \tag{5}$$

For the different inspection intervals, the $F_i(x)|^{k\tau}_{(k-1)\tau}$ indicates the probability of the soft component i failure at $((k-1)\tau, k\tau]$ interval, when the soft component is on safety condition in the last interval. As well, $F_i(x)|^{\tau}_{0}$ indicates the probability of the soft component i failure at $(0, \tau]$ interval, when the soft component i is on unsafety condition in the last interval (note, the perfect repair just done in inspection). According to Eqs. (5) and (3), $p^{k}_{\tau,i}(t)$ can be simplified as follows:

$$\begin{aligned} p^{k}_{\tau,i}(t) = p^{k-1}_{\tau,i}(t)\left[e^{-\left[\left(\frac{k\tau}{\theta_i}\right)^{\beta_i} - \left(\frac{(k-1)\tau}{\theta_i}\right)^{\beta_i}\right]}\right] + (1 - p^{k-1}_{\tau,i}(t))\left[e^{-\left[\left(\frac{\tau}{\theta_i}\right)^{\beta_i}\right]}\right], k \\ = 1, \ldots, T/\tau \end{aligned} \tag{6}$$

For example, from two inspection frequency

$$p^{1}_{T/2,i}(t) = 1 \times \left[e^{-\left[\left(\frac{T/2}{\theta_i}\right)^{\beta_i} - \left((1-1)\left(\frac{T/2}{\theta_i}\right)\right)^{\beta_i}\right]}\right] + (1-1) \times [\ldots] = e^{-\left(\frac{T/2}{\theta_i}\right)^{\beta_i}} \tag{7}$$

$$p^{2}_{T/2,i}(t) = p^{1}_{T/2,i}(t)\left[e^{-\left[\left(\frac{2T/2}{\theta_i}\right)^{\beta_i} - \left((2-1)\left(\frac{T/2}{\theta_i}\right)\right)^{\beta_i}\right]}\right] + (1 - p^{1}_{T/2,i}(t)e^{-\left(\frac{T/2}{\theta_i}\right)^{\beta_i}} \tag{8}$$

The cycle T is the planning horizon (e.g. 1 year) which is fixed. In the cycle T, the soft component is inspected at times, $k\tau$ (k = 1, 2, …, n), where $T = n\tau$. Failures of the soft component are perfectly repaired if failure accurse. We assume that inspection and possible repairs are also done at the end of the cycle T (last inspection is on the end of cycle T), that is, for $k = n$. The objective is to find the optimal risk based inspection interval that can minimize the expected total cost of the soft component incurred over the cycle T. When the soft component fails, it remains in a failed state until the next inspection time. Therefore, if the soft component failed in each inspection interval, a downtime penalty cost is incurred. The cost is proportional to the elapsed time from failure time to its detection at inspection time. Thus, the costs for resulting from the soft component in each of the inspections k, k = 1,2, …, n includes the cost of inspection for component i, C^{s}_{i}, the cost of repair if found fails for component i, C^{d}_{i}, and the penalty cost for the elapsed time for the failure of component i, C_i^p, thus, the expected cost incurred in the inspection k in the cycle T is proposed by Rezaei and Imani [3] and extended for multi component as follow:

$$\underset{\forall\tau=T,T/2,\ldots,1}{E[C^{\tau}]} = \sum_{i=1}^{m}\sum_{k=1}^{T/\tau} E\left[C_i^{(k-1)\tau,k\tau}\right] = \left(\begin{array}{c} \displaystyle\sum_{i=1}^{m}\sum_{k=1}^{T/\tau} C_i^s + \sum_{i=1}^{m}\sum_{k=1}^{T/\tau} C_i^d\left[1-p_{\tau,i}^k(t)\right] + \\ \displaystyle\sum_{i=1}^{m}\sum_{k=1}^{T/\tau} C_i^p\left[\tau\left(1-p_{\tau,i}^k(t)\right)\right] \end{array}\right)$$
$$= \left(\sum_{i=1}^{m}\left({}^{T}\!/_{\tau}\right)C_i^s + \left(C_i^d+\tau C_i^p\right)\left[\sum_{i=1}^{m} {}^{T}\!/_{\tau} - \sum_{i=1}^{m}\sum_{k=1}^{T/\tau} p_{\tau,i}^k(t)\right]\right) \tag{9}$$

In common inspection, the $\sum_{i=1}^{m}\sum_{k=1}^{T/\tau} C_i^s$ replaces with $\sum_{k=1}^{T/\tau} C_{common}^s$.

For risk consideration, there are no accurate mathematical formula for considering risk in inspection. In reality, considering risk is based on expert's judgment. For short planning horizon length (T), the $\left(1+\tilde{\psi}\tau\right)\tilde{E}[C^T]$ risk statement is useful, where, $\tilde{\psi}$ is fuzzy risk present [3]. In addition, applying the $\exp(\tau\tilde{\psi})\tilde{E}[C^T]$ in long planning horizon length is useful. But, these equations are not general or permanent. So, applying the expert judgment can be more suitable to considering risk and reduce inspection interval time. The fuzzy risk scale proposed by Rezaei and Imanni [3] is presented as Table (2).

To gain optimal risk based inspection interval time for $\tilde{\psi} = (a, b, c)$ the Eq. (10) is proposed as follow.

$$\begin{aligned} \tilde{\tau}^* &= \tau - \tilde{\psi}\tau = \tau - (a,b,c)\tau = (\tau - a\tau, \tau - b\tau, \tau - c\tau) \\ &= \frac{(\tau - a\tau) + 4\times(\tau - b\tau) + (\tau - c\tau)}{6} \end{aligned} \tag{10}$$

3 Numerical Example

Let us consider a general infusion pump adopted from a case study reported in [22]. The infusion pump is used to accurately deliver liquids through intravenous or epidural routes for therapeutic and/or diagnostic purposes. Here, we have assumed same values for the components' failure rate parameters as estimated in [22] (Table 3).

The proposed model is codes by software. The expected total cost is calculated by the proposed model in which MATLAB software (version 2015) is employed to increase the correctness of calculation. To indicating reliability analysis, the

Table 2 Risk scale

Average	(0.45,0.50,0.55)	Negligible	(0,0.05,0.10)
Between average and relatively strong	(0.50,0.55,0.60)	Very very low	(0.05,0.10,0.15)
Relatively strong	(0.55,0.60,0.65)	Between very very low and very low	(0.10,0.15,0.20)
Between relatively strong and strong	(0.60,0.65,0.70)	Very low	(0.15,0.20,0.25)
Strong	(0.65,0.70,0.75)	Between very low and low	(0.20,0.25,0.30)
Between strong and very strong	(0.70,0.75,0.80)	Low	(0.25,0.30,0.35)
Very strong	(0.75,0.80,0.85)	Between low and relatively low	(0.30,0.35,0.40)
Between very strong and very very strong	(0.80,0.85,0.90)	Relatively low	(0.35,0.40,0.45)
Very very strong	(0.85,0.90,0.95)	Between relatively low and average	(0.50,0.45,0.50)

summary of $p_{\tau,1}^{k}(t)$ results for component 1 is presented. The $p_{\tau,1}^{k}(t)$ values for component 1 is presented in Table (4). The results from Table (4) indicate the improvement of $p_{\tau,1}^{k}(t)$ with decreasing inspection interval time (τ) and fixed inspection number (k) (the value of each column is on increasing). As well, for each fixed inspection interval time (τ) and increasing inspection number (k), the $p_{\tau,1}^{k}(t)$ it's on decreasing.

Increasing the number of inspections increases the inspection costs and reduces the downtime penalty cost. The contrast between these two costs caused the Non-strict total cost plot. In riskless model, the optimal inspection interval obtain for 6 inspection frequencies and it's related to $\tau = 2$.

From Eq. (9), total expected cost for each fixed inspection interval time (τ) with inspection cost, repair cost, and downtime penalty cost are shown in Fig. (1). The plots of inspection and repair cost have ascending trend. Contrary of them, the downtime penalty cost has decreasing trend.

As mentioned, risk reduces inspection interval time. In Table 5, the total expected cost for different present of risk is presented. As indicated, with increasing risk ($\tilde{\psi}$) the optimal inspection interval is reduces. From Eq. (10):

Table 3 Failure rate functions' parameters, different costs, planning horizons corresponding to components 1–5

Component i	β	θ	C_i^s	C_i^d	C_i^p	T
1	1.3	3.5	40	70	100	12
2	1.1	4.6	40	45	25	12
3	2.1	6	40	100	200	12
4	1.8	10	40	75	50	12
5	1.7	3.6	40	150	150	12

Table 4 The $p^{k}_{\tau,1}(t)$ results for $\tau = 12, 6, 4, \ldots, 1$ and different k

Inspection intervals	Sub-inspection intervals				
	K = 1	K = 2	K = 3	K = 4	K = 5
τ = 12	0.007				
τ = 6	0.1333	0.1225			
τ = 4	0.304353	0.2652	0.2584		
τ = 3	0.441134	0.3798	0.3674	0.3572	
τ = 2.4	0.542086	0.4696	0.4524	0.4389	0.43

4 Conclusion

The inspections prevent failures and decrease maintenance cost. The systems have been more reliable by inspection activity. In the proposed model, the expected total cost associated to the soft components has been formulated to finding optimal inspection interval. Then, the expected total cost is evaluated for different number of inspections in a cycle to determine the optimal value of inspection interval. Risk have critical role in maintenance and inspection interval. Results show that risky equipment has shorter inspection interval than riskless equipment. For example, a transformer with few failure will have short inspection intervals time because of risk existence. So, the risk based optimization inspection interval is so applicable in industries. In this study, the fuzzy number is used to model uncertainty in expert judgment to consider risks. As indicated in Table 4, with increasing risk ($\tilde{\psi}$) the optimal inspection interval reduced. The other characteristic of this paper, is to

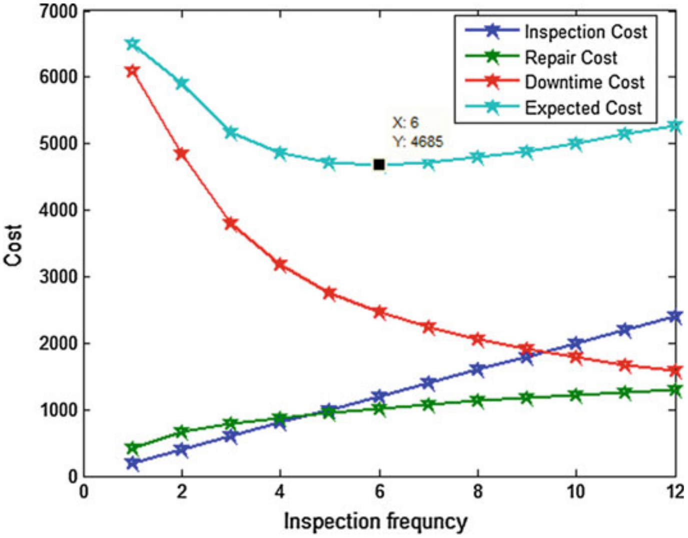


Fig. 1 The riskless costs resulting for different inspection frequencies τ = 12, 6,…, 1

Table 5 Risk analysis

Risk	Optimal risk based inspection interval
$\tilde{\psi} = (0, 0.05, 0.1)T$	$\tilde{\tau}^* = \frac{(2) + 4\times(2-0.05\times2) + (2-0.1\times2)}{2} = 1.95$
$\tilde{\psi} = (0.05, 0.1, 0.15)T$	$\tilde{\tau}^* = \frac{(2-0.05\times2) + 4\times(2-0.1\times2) + (2-0.15\times2)}{6} = 1.9$
$\tilde{\psi} = (0.1, 0.15, 0.2)T$	$\tilde{\tau}^* = \frac{(2-0.1\times2) + 4\times(2-0.15\times2) + (2-0.2\times2)}{6} = 1.85$
$\tilde{\psi} = (0.15, 0.2, 0.25)T$	$\tilde{\tau}^* = \frac{(2-0.15\times2) + 4\times(2-0.2\times2) + (2-0.25\times2)}{6} = 1.8$
$\tilde{\psi} = (0.2, 0.25, 0.3)T$	$\tilde{\tau}^* = \frac{(2-0.2\times2) + 4\times(2-0.25\times2) + (2-0.3\times2)}{6} = 1.75$
$\tilde{\psi} = (0.25, 0.3, 0.35)T$	$\tilde{\tau}^* = \frac{(2-0.25\times2) + 4\times(2-0.3\times2) + (2-0.35\times2)}{6} = 1.7$
$\tilde{\psi} = (0.3, 0.35, 0.4)T$	$\tilde{\tau}^* = \frac{(2-0.3\times2) + 4\times(2-0.35\times2) + (2-0.4\times2)}{6} = 1.65$
$\tilde{\psi} = (0.35, 0.4, 0.45)T$	$\tilde{\tau}^* = \frac{(2-0.35\times2) + 4\times(2-0.4\times2) + (2-0.45\times2)}{6} = 1.6$

prefer grouping maintenance to individual maintenance as economic failure dependency. The model is so applicable for soft system with multi repairable component such as turbine rotor, transformer….

References

1. Rezaei E, Imani DM (2015), A new modeling of maintenance risk based inspectioninterval optimization with fuzzy failure interaction for two-component repairable system. 6(31)
2. Rezaei E, Imani DM (2015) A New Modeling of Periodic Inspection Interval Optimization in Computer System with Failure Interaction. Australasian Journal of Computer Science. doi:10.3923/aujcs
3. Rezaei E, Imani DM (2015) A new modeling of maintenance risk based inspection interval optimization with fuzzy failure interaction for two-component repairable system. IJONS/PUB-31/VOL-5/Aug/01 01/2015
4. Chen S et al (2014) A bi-objective maintenance scheduling for power feeding substations in electrified railways. Transp Res Part C 44(2014):350–362
5. Wintle JB (2001) Best practice for risk based inspection as a part of plant integrity management. Health and Safety Executive (HSE), London
6. Elaya K (2014) Perumal, corrosion risk analysis, risk based inspection and a case study concerning a condensate pipeline. Procedia Engineering 86:597–605
7. Sarkar J, Sarkar S (2000) Availability of a periodically inspected system under perfect repair. J Stat Plann Infer 91:77–90
8. Barlow RE, Hunter LC, Proschan F (1963) Optimum checking procedures. J Soc Ind Appl Math 11(4):1078–1095
9. Nakagawa T (2005) Maintenance theory of reliability. Chapter 8. Springer: London
10. Taghipour S, Banjevic D (2011) Periodic inspection optimization models for a repairable system subject to hidden failures. IEEE Trans Reliab 60(1):275–284

11. Ahmadi A, Kumar U (2011) Cost based risk analysis to identify inspection and restoration intervals of hidden failures subject to aging. IEEE Trans Reliab 60(1):197–209
12. Thomas LC, Jacobs PA, Gaver DP (1987) Optimal inspection policies for standby system. Commun Stat Stoch Models 3:259–273
13. Wang H (2002) A survey of maintenance policies of deteriorating systems. Eur J Oper Res 139(3):469–489
14. Cho DI, Parlar M (1991) A survey of maintenance models for multi-unit systems. Eur J Oper Res 51(1):1–23
15. Sarkar A, Panja SC, Sarkar B (2011) Survey of maintenance policies for the last 50 Years. Int J Softw Eng Appl 2(3):130
16. Nicolai RP, Dekker R (2007) Optimal maintenance of multi-component systems: a review. In: Murthy DNP and Kobbacy KAH (eds) Complex system maintenance handbook. Springer, ch. 11, pp 263–286
17. Murthy DNP, Nguyen DG (1985) Study of a multi-component system with failure interaction. Eur J Oper Res 21(3):330–338
18. Murthy DNP, Nguyen DG (1985) Study of a two-component system with failure interaction. Naval Res Logist 32(2):239–247
19. Scarf PA, Deara M (1998) On the development and application of maintenance policies for a twocomponent system with failure dependence. IMA J Math Appl Bus Ind 9:91–107
20. Scarf PA, Deara M (2003) Block replacement policies for a two-component system with failure dependence. Naval Res Logist 50(1):70–87
21. Zequeira RI, Berenguer C (2004) Maintenance cost analysis of a two-component parallel system with failure interaction. In: Proceedings reliability and maintainability symposium 2004, pp 220–225
22. Taghipour S, Banjevic D, Jardine AKS (2010) Periodic inspection optimization model for a complex repairable system. Reliab Eng Syst Saf 95(9):944–952
23. GolMakani HR, Moakedi H (2012a) Periodic inspection optimization model for a two-component repairable system with failure interaction. Comput Ind Eng 63:540–545
24. Hamid Reza Gol Makani and Hamid Moakedi (2012) Periodic inspection optimization model for a multi-component repairable system with failure interaction. Int J Adv Manuf Technol 61:295–302

Part VII
Reliability and Maintenance of Mining Machinery

Reliability Analysis and Maintenance Scheduling of the Electrical System of Rotary Drilling Machines

Mohammad Javad Rahimdel, Mohammad Ataei and Reza Khalokakaei

Abstract This paper studied reliability of electrical system for electrical-hydraulic rotary drilling machines. Reliability is the most important characterization of repairable systems. Electrical systems have a vital role in any electrical-diesel or electrical-hydraulic machines. As any failure in electrical system finally leads to stopping the machine, research on the reliability of this system is essential. In this research reliability of electrical system of drilling machines in Sarcheshmeh copper mine in Iran had been modelled and analysed. There were four hydraulic-electric machines in this mine (named A, B, C and D), that all of them had been selected for data collection and analysis of failure. The results of statistical analysis showed that time between failures (TBF) data of this system follows the gamma distribution for machines A and C and weibull (3P) and exponential distributions for machines B and D, respectively. Also, the results showed that with considering 90 % for preventive maintenance (PM) interval, after the first maintenance the reliability of this system will be improved by 5.23, 7.22, 3.62 and 5.26 % respectively for machines A, B, C and D.

Keywords Rotary drilling machines · Electrical system · Reliability · Preventive maintenance

M.J. Rahimdel (✉)
Department of Mining Engineering, Sahand University of Technology, Tabriz, Iran
e-mail: m_rahimdel@sut.ac.ir

M. Ataei · R. Khalokakaei
Faculty of Mining Engineering, Petroleum and Geophysics,
Shahrood University of Technology Shahrood, Shahrud, Iran
e-mail: ataei@shahroodut.ac.ir

R. Khalokakaei
e-mail: r_kakaei@shahroodut.ac.ir

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_45

1 Introduction

Drilling is the first step of exploitation process in large surface mining. Nowadays, rotary blasthole drilling in almost 98 % of big open pit mines and quarry is the most repute method to rock penetrating and drilling. Reliability is one of the most important performance measures for repairable system designers and operators [1]. As yet, many researches on reliability of more engineering systems have been done; however, the researches on drilling blasthole machines have been based on empirical methods and engineering judgments. Regarding to all of the main components of drilling machines such as hydraulic pumps and motors, compressor, filters, electrical sensors, heaters, coolers etc. work electrically, so as result any failure of electrical system will stop the machine, eventually. Thus, this paper has focused on reliability modelling and analysing of electrical system of drilling machines. For this reason, drilling machines of Iran's Sarcheshmeh Copper Mine have been selected to data collection and analysis. There are four rotary drilling machines in this mine, named machines A, B, C and D, that all of them are hydraulic-electric type and the source and distribution of power are electrical and hydraulic, respectively. Therefore, the electrical system has an important and vital role in this type of drilling machines. The main components of these machines are explanted as follow [2–4]:

(1) Main Electrical motor
This motor is the main source of power. All the driven components in the machine are driven by use of this power source so that they generate desired movements of the components. Electrical current is supplied from main electrical network of mine.

(2) Starter motor
Starter motor is a small electrical motor that used for running the main electrical motor.

(3) Cable Reel Electrical Motor
A cable reel, mounted on the front end of a drilling blasthole machine. A cable reel through an electric motor automatically ensures tidily and tightly wound position of the power cable that supplies power from the mine power supply to the machine even as the machine moves from one blasthole to another.

(4) Auxiliary winch
Almost all the accessories used in rotary drilling blasthole machine are so heavy that they cannot be manually lifted, shifted and handled. Therefore, almost every rotary drilling blasthole machine is provided with a wire rope and an auxiliary winch. In Sarcheshmeh copper mine drilling machines, the winch is powered by electric motors through a planetary reduction gear box for compactness. The winch placed near the lower end of the mast.

(5) Heaters
Heaters become essential when the machine has to operate in cold weather. Apart from the operator's cab and machinery house, heaters have to be fitted on the rotary head gear case and the cases of gears that reduce the speed of the

propel motors. In this case, heaters are heating coil type. Water tank and hydraulic oil tank have this heater type. In most cases such as operator cap and machinery house, a fan is provided near the heater so it spreads the hot air within all the space of the enclosures.

(6) Sensors and Gauges
The console in front of the operator cab contains many indicators and controls, such as oil level, fuel level, water injection tank level and drill level indicators voltmeter or ammeter and engine hour meter, engine water temperature, compressor temperature, engine hour meter and so on gages. Also, these machines equipped with electronically operated depth indicators that indicate blasthole depth inside cab after sensing the drill pipe addition. Drill level or depth indicator one of the most important of this sensors which numerically shows the blasthole depth.

2 Background

2.1 Reliability Analysis

Reliability is the probability that an item will perform its assigned mission satisfactorily for the stated time period when used according to the specified conditions. The basic reliability function is defined by Eq. (1) [1].

$$R(t) = 1 - F(t) = 1 - \int_0^t f(x)dx \tag{1}$$

Where $R(t)$ is the reliability at time t; $F(t)$ is cumulative failure distribution function and $f(x)$ is failure probability density function. The reliability characteristic of a piece of equipment can be determined, by analysing of the time between failures (TBF) data. For the reliability modelling of repairable systems, the basic methodology step-by-step is presented in Fig. 1 [5]. It shows a detailed flowchart for model identification and is used here as a basis for the analysis of the failure data.

Three methods are generally used for reliability analysis of repairable systems including Renewal Process (RP), Homogeneous Poisson Process (HPP), and Non-Homogeneous Poisson Process (NHPP). In RP method, analysis of data reliability is usually based on the assumption that the times between failures are independent and identically distributed (iid) at the time domain. Trend test and serial correlation test are used for validation of this assumption.

The trend test involves plotting the cumulative failure numbers against the cumulative time. If one obtains a curve that is approximately a straight line, then the data is identically distributed and free from trends. The data sets can also be analysed for the presence of trends by using the test suggested in military hand book-189 by calculating the test statistic as follows [6]:

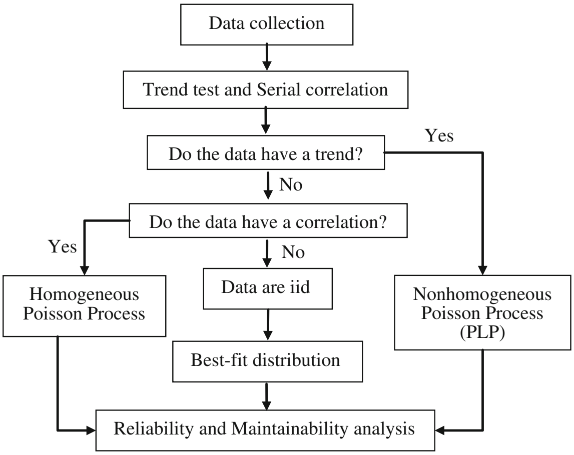

Data collection
Trend test and Serial correlation
Do the data have a trend?
Yes
No
Do the data have a correlation?
Yes
No
Homogeneous Poisson Process
Data are iid
Nonhomogeneous Poisson Process (PLP)
Best-fit distribution
Reliability and Maintainability analysis

Fig. 1 Reliability analysis process of a repairable system [5]

$$U = 2\sum_{i=1}^{n-1} \ln(T_n/T_i) \tag{2}$$

where, the data are failure-truncated at the nth failure at time T_n.

Under the null hypothesis of a homogeneous Poisson process, the test statistic U is chi-squared distributed with a 2(n-1) degree of freedom. The presence of serial correlation can be tested by plotting the ith TBF against (i-1)th TBF. If the plotted points are randomly scattered without any pattern, it can be interpreted that the TBFs are free from serial correlation.

2.2 Preventive Maintenance and Reliability Improvement

To keep a system in normal condition, taking proper maintenance becomes even more important during its serviced life. Maintenance was classified into two categories, corrective maintenance (CM) and preventive maintenance (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.

Preventive maintenance is predetermined work performed to a schedule with the aim of preventing the wear and tear or sudden failure of equipment components. PM helps to:

- Protect assets and prolong the useful life of production equipment.
- Improve system reliability.
- Decrease cost of replacement.
- Decreases system downtime.
- Reduce injury.

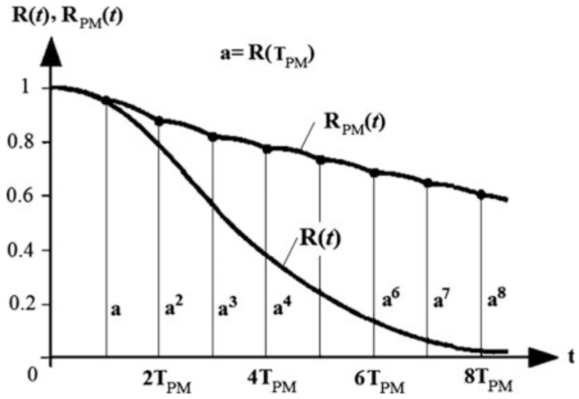


Fig. 2 Typical effect of PM on reliability functions of a system with maintenance interval T_{PM} [8]

Reliability Cantered Maintenance (RCM) methodology offers the best available strategy for PM optimization. RCM is definition as a method for developing and selection maintenance design alternative based on safety, operational and economic criteria, RCM employs a system perspective in its analysed of system functions, failures of functions, and preventive of these functions [7].

In this paper, after a reliability modelling, a critical allowed level of reliability has been defined for electrical system of each machine, which means that the system continued to operation in this reliability level without falling below this level of reliability. Based on this critical level, PM intervals can be defined. Figure 2 shows the results of reliability improvement exactly after the first PM operation. The reliability function after PM ($R_{PM}(t)$) can be calculated by followed Eq. (3) [8]:

$$R_{PM}(t) = \begin{cases} R(t), 0 < t \leq T_{PM} \\ R^n(T_{PM})R(t - nT_{PM}), \\ nT_{PM} \leq t < (n+1)T_{PM}, n \geq 1 \end{cases} \tag{3}$$

where $R(t)$ is the reliability of failure-free time t; $R_{\mathrm{PM}}(t)$ is the reliability function after preventive maintenance; T_{PM} is the preventive maintenance intervals and n is the number of preventive maintenance which have been done.

3 Reliability Analysis: A Case Study

3.1 Failure Data Collection and Analysis

In this research, all of four drilling machines in Sarcheshmeh copper mine are selected to data collection and analysis. Then, the TBF data of their electrical system had been calculated over a period of 18 months [9–10]. The results of the

Table 1 Computed values of the test statistic U for TBF

Machine	Number of failure	Degree of freedom	Calculated statistic U	Rejection of null hypothesis at 5 % level of significance	Modelling method
A	81	160	202.78	Not rejected (>61.28)	Renewal process
B	82	162	152.45	Not rejected (>62.16)	Renewal process
C	34	66	55.19	Not rejected (>21.82)	Renewal process
D	40	78	85.48	Not rejected (>26.50)	Renewal process

trend and serial correlation tests on available data of four machines were shown approximately a straight line. Thus, the data were free of trend. The plotted points in serial correlation tests were randomly scattered without any pattern, therefore, the data were free from serial correlations, too.

The data set was also analysed for the presence of trend by using the MIL-HDBK-189 Test. The computed values of the test statistic (Eq. 2) for available TBF data are given in Table 1.

3.2 Data Analysis and Reliability Modelling

Both graphical and analytical methods show that the data are free of trend and serial correlation. As a result, renewal process techniques can be used for reliability modelling. The reliability of electrical system was calculated by the use of best-fitted distribution.

Data analysis and finding the best-fit distributions were done with using *Easyfit* 5.5 software. The Kolmogorov-Smirnov (K-S) test has been used for selecting the best distributions for reliability analysis. The results of data analysis and the best-fitted distributions are illustrated in Table 2.

Regarding to Table 2, the achieved reliability plots have been shown in Fig. 3.

Regarding to Fig. 3, the reliability of the electrical system reduces to zero after 500 h operation of machines A and B and 1400 h operation of machines C and D. On the other hand, after 500 h operation of machines, only two machines will be active, namely machines C and D, and 900 h later all of four machines will be failed, due to full failing of their electrical systems. Electrical systems of machines A and B have a similar reductive behaviour in reliability and in comparison of machines A and B have a faster reductive rate. Such that only after 8 h operation (or at the end of first shift operation) reliability of this system in machine A and B reduced by 20 %. At this time, electrical system reliability of machines C and D is reduced by only 5 and 1 %. These two machines have a similar electrical system

Table 2 Results of the best fitted distributions

Machine	A	B	C	D
Distribution	K-S test	K-S test	K-S test	K-S test
Exponential	0.162	0.1818	0.1359	0.0827
Weibull (3P)	0.0965	0.0628	0.0954	0.1147
Gen. gamma	0.0676	0.1226	0.078	0.0993
Gamma	0.0559	0.1417	0.0717	0.0926
Weibull (2P)	0.0683	0.0694	0.105	0.0891
Best distribution	Gamma	Weibull (3P)	Gamma	Exponential
Parameters	α = 0.6	α = 0.698	α = 0.775	λ = 0.0045
		β = 66.66		
	β = 120.957	γ = 0.125	β = 396.47	γ = 11.125

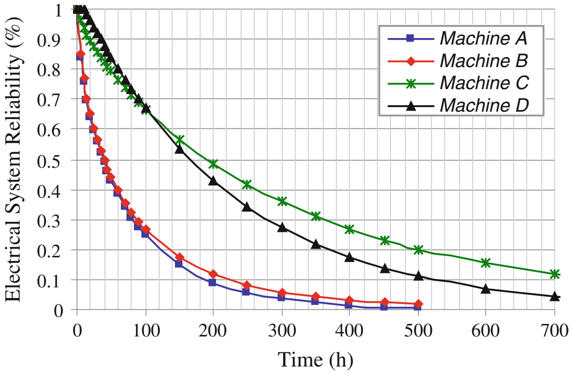


Fig. 3 Reliability plots of electrical system of drilling machines

reliability, but in spite of their lesser with lesser reductive rate. Machines A and B have passed about 16 years of operation, on the other hand the machines C and D are in about 10th year of operation. So, the lowest reliability of these machines is related to oldness of machines and the needs of all parts to fundamental repair or replace. As it can be seen from reliability plots, machines A and B have a higher reliable electrical system rather than the others at the all of operation time. After 250 h operation, reliability of electrical system of machines A and B will be reduced to lower 10 % and stopped at 700 h operation. From this time onward, there are only two active machines (C and D) and with very low reliability level (lower than 10 %). Also after 1400 h from starting to drilling operation, both of machines will be stopped. Without considering any preventive maintenance for electrical system before 1400 h operation, drilling fleet of mine will be stopped.

3.3 Maintenance Scheduling and Reliability Improvement

In many engineering operations, 80 % is selected as the best practical value for performance evaluation. Because of the importance and vital role of electrical system in drilling machines, in this paper the 90 % is selected as reliability level for scheduling the preventive maintenance. Consequently, it has been suggested that the preventive maintenance should be done every 2.5, 5, 15 and 30 h, respectively for machines A, B, C and machine D. On the other words, TPM for these machines will be equal to 2.5, 5, 15 and 30 h, respectively. Regarding to Eq. (3) and TPM of machines' electrical system, RPM has been calculated for different reliability levels which have been shown in Fig. 5.

Regarding to Fig. 4, with carrying out the preventive maintenances on electrical system of machines, their reliability will be improved. There is an increase of 0.1–6.54 %, 0.1–7.25 %, 0.1–3.73 % notably in electrical system reliability of machines A, B and C. Nevertheless, it should be noted that remarkably after 10, 30 and 125 h the plot of reliability "with PM" and "without PM" of these machines meets to each other. This means that after this period of times the short and fast preventive services cannot compensate the failures of system. Therefore, the electrical system of these machines should be fundamentally serviced and maintained.

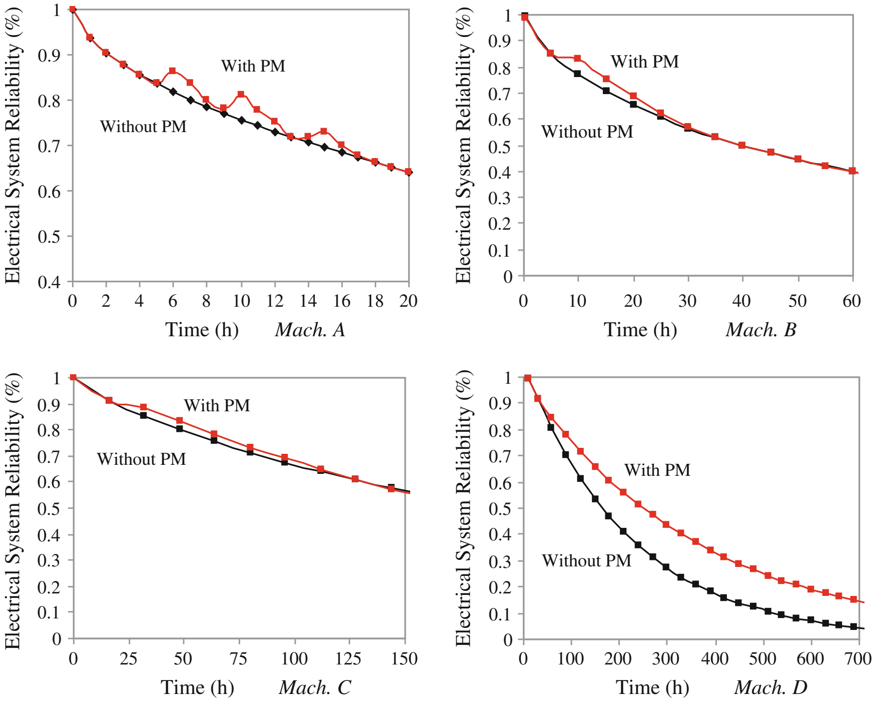


Fig. 4 Effect of preventive maintenance on reliability of electrical system of drilling machines

But by considering the preventive maintenance, the reliability plot of machine D shows that the reliability of electrical system will be increased by 5.31 % only 30 h after the first PM, that is after second PM (with TPM = 30 h). This reliability level at the third PM or after 90 h operation will be increased by 10.25 %. These calculations show that the reliability of electrical system of machine D with PM has a higher and longer increasing rate compared to other machines.

4 Conclusions

In this research, the reliability of electrical system of all drilling machines in Sarcheshmeh copper mine in Iran were modelled and analysed. The main results of this research can be summarized as following:

- The assumption that the failure data of transmission subsystem is trend free is valid for all machines. The serial correlation test showed that the data are correlation free and, as a result, the data of the machines are independent and identically distributed (iid).
- Analysis showed that the time between failures (TBF) of machines A and C obey the gamma distribution. Also, TBF of machines B and D obey the weibull (3P) and exponential distributions, respectively.
- The reliability plot of machines A and B is similar to each other and reliability of electrical system of these two machines are in highest than all other machines and reached to zero after 500 h. Also, after 1400 h from starting to operation of machines, drilling fleet of mine will be completely stopped.
- After 10, 30 and 125 h operation of machines A, B and C, respectively, the plot of reliability “with PM” and “without PM” of these machines meets to each other. Nevertheless, due to effects of the preventive maintenance on reliability, if the electrical system of machines A and B to be checked and serviced every 10 (or at the second PM interval), the reliability of this system of machines will be improved by 5.21, 7.25 %, respectively. Also, with considering the 30 and 60 h as PM intervals for machines C and D, the reliability of electrical system of these machines will be noticeably increased by 3.73 and 5.31 %.

Acknowledgments The authors thank the R & D centre of Iranian National Copper Company for its financial support. The kind help of the drilling office and workshop of Sarcheshmeh copper mine is also acknowledged.

References

1. Dhillon BS (2008) Mining equipment, reliability, maintainability and safety. Springer Series in Reliability Engineering, pp 27–38
2. Gokhale BV (2011) Rotary drilling and blasting in large surface mines. CRC Press-Balkema, Boca Raton, pp 153–222

3. Jimeno CL, Jimeno EL, Carcedo FJA (1995) Drilling and blasting of rock. A.A. Balkema, Rotterdam, pp 48–52
4. Ingersoll-Rand (1990) Airend overhaul manual for models. Rock Drill Division
5. Ascher H, Feingold H (1984) Repairable system reliability. Dekker, New York
6. Kumar U, Klefsjö B (1992) Reliability analysis of hydraulic system of LHD machine using the power low process model. Reliab Eng Syst Saf 35:217–224
7. Jones RB (1997) Risk based management. Jaico Publication house, India
8. Kececiyoglu DB (2002) Reliability Engineering Handbook. DEStech Publication, Lancaster, vol 1, p 721
9. Rahimdel MJ, Ataei M, Khalokakaei R, Hoseinie SH (2013) Reliability-based maintenance scheduling of hydraulic system of rotary drilling machines. Int J Min Sci Technol 23 (5):771–775. doi:10.1016/j.ijmst.2013.08.023
10. Rahimdel MJ, Ataei M, Kakaei R, Hoseinie SH (2013) Reliability analysis of drilling operation in open pit mines. Arch Min Sci 58(2):569–578. doi:10.2478/amsc-2013-0039

Monte Carlo Reliability Simulation of Underground Mining Drilling Rig

Hussan Al-Chalabi, Hadi Hoseinie and Jan Lundberg

Abstract Drilling rigs are widely used in mine development or construction and tunnel engineering projects. The rig consists of 12 subsystems in a series configuration and can be driven by diesel or electrical engines. This paper uses the Kamat-Riley (K-R) event-based Monte Carlo simulation method to perform reliability analysis of an underground mine drilling rig. For data analysis and to increase statistical accuracy, the paper discusses three case studies in an underground mine in Sweden. Researchers built a process to programme the simulation process and used MATLAB$^{\text{TM}}$ software to run simulations. The results showed the simulation approach is applicable to the reliability analysis of this rig. Moreover, the reliability of all rigs reaches almost zero value after 50 h of operation. Finally, the differences between the reliability of the studied fleet of drilling rigs are a maximum 10 %. Therefore, all maintenance or spare part planning issues can be managed in a similar way.

Keywords Drilling rigs · K-R method · Monte carlo simulation · Reliability · Simulation · Underground mines

H. Al-Chalabi (✉) · H. Hoseinie · J. Lundberg
Division of Operation and Maintenance Engineering, Luleå University of Technology,
971 87 Luleå, Sweden
e-mail: hussan.hamodi@ltu.se

H. Hoseinie
e-mail: hadi.hoseinie@ltu.se

J. Lundberg
e-mail: Jan.Lundberg@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_46

1 Introduction

Underground mines are a main source of minerals. The growing demand for metals as a result of modern lifestyles and ongoing industrial development has focused attention on factors affecting the extraction of minerals. One of the most important factors is the unscheduled stoppage of machines used in the extraction of ore [1]. Economic globalisation is increasing competition among mining companies, pushing them to achieve higher production rates by increasing automation and mechanisation and using new and more effective equipment. This forces companies to buy more reliable capital equipment with higher performance capabilities; naturally, these are more expensive. At the same time, the equipment used in underground mining industries is subject to degradation throughout its operating life; this increases the operating and maintenance costs and reduces production rates, causing a negative economic effect as equipment ages [2].

The drilling rig is very important to the extraction process. At its most basic level, drilling is the process of making holes in the mining room face, but reliable and accurate drilling operation facilitates the rest of the production chain and improves the economic and safety issues of the mine. All drilling machines for mining applications are composed of similar operational design units, including cabin, boom, rock drill, feeder, service platform, front jacks, hydraulic pump, rear jack, electric cabinet, hose reeling unit, cable reeling unit, diesel engine, hydraulic oil reservoir, operator panel and water tank. Drilling rigs manufactured by different companies have different technical characteristics, e.g. capacity and power. Based on the operating manuals, field observations and maintenance reports from the collaborating mine, in this study, the drilling rig is considered a system divided into several subsystems and connected in series configuration; if any subsystem fails, the operator will stop the rig to maintain it. Given this configuration, having good knowledge about these rigs and properly maintaining them is essential for a reliable drilling operation and assured production.

Collecting data, analysing data and making decisions are time consuming process, but they should be done during any reliability study. The reliability analysis of mining machines is especially difficult in practice because of the special operation and maintenance environment and the work pressure in mines [3].

This paper uses stochastic simulation to evaluate the reliability of three drilling machines used in an underground mine in Sweden. Stochastic simulation is a suitable technique to assess the reliability of a system and can be applied in two ways [4, 5]:

- Sequential approach by examining each basic interval of the simulated period in chronological order, and
- Random approach by examining randomly chosen basic intervals of the system's lifetime.

The second approach, usually known as "Monte Carlo" method, is selected for this paper. This is a numerical method which allows the solution of mathematical and technical problems by means of system probabilistic models and simulation of random variables.

Many researchers have studied the reliability and maintainability of mining equipment and its failure behaviour. For example, [6] analysed the operational reliability of a fleet of diesel operated load-haul dump (LHD) machines in Kiruna mine in Sweden [6]. Later, [7] performed reliability analysis on the power transmission cables of electric mine loaders in Sweden [7]. Reliability assessment of mining equipment was performed by [8]; using genetic algorithms, they developed and tested mobile mining equipment reliability assessment models [8]. Vayenas and Xiangxi [9] studied the availability of 13 LHD machines in an underground mine. They were interested in the influence of machine downtime on productivity and operation costs and used a reliability-based approach and a basic maintenance approach to determine the machine's availability [9]. More recently, [10] used fault tree analysis (FTA) to analyse the idle times of automated load-haul-dump LHD machines at a Swedish underground mine [10]. Finally, [4] performed reliability modelling on the drum Shearer machine used at Taba's coal mine in the central desert of Iran and analysed the failure rate of the machine's subsystems [11].

Although there are many reliability and maintainability studies of underground mining equipment, no one has looked specifically at drilling machines. Given the importance of underground mining mobile equipment for production, not to mention the complexity of the equipment and the harsh mining environment, reliability analysis of the drilling rig must meet rigorous requirements. Thus, the aim of this paper is to analyse and compare the reliability of several drilling rigs to show the Kamat-Riley (K-R) event-based Monte Carlo simulation method can be used to simulate the reliability of repairable complex systems based on available data from the case study mining company. The paper also aims to shed light on the reliability behaviour of the mining drilling rig to enhance decision-making, improve reliability and reduce downtime.

2 Mining Drilling Rig

A mining drilling rig is used to dig holes in the ground. For example, mobile drilling rigs can be used to make tunnels and underground facilities, and small or medium-sized mobile drilling rigs are appropriate for mineral exploration. Mining drilling rigs are used for two main purposes, namely production drilling (for processes in the mining production cycle such as bolting) and exploration drilling (to identify the location of minerals). Figure 1 illustrates the process cycle for drift mining; as the figure shows, drilling is a key step. From an economic viewpoint, drilling rigs make an important contribution to the mine's production rate but they have a high acquisition, maintenance and operating cost and represent a possible critical bottleneck for production [12].

Economic competition has pressured mining companies into achieving higher production rates by enhancing the techniques of drilling and blasting and increasing

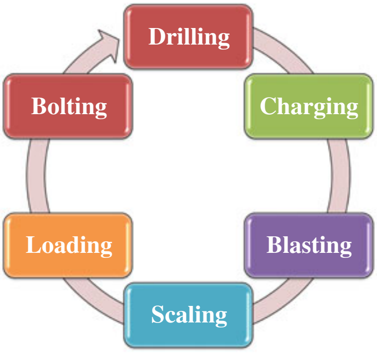


Fig. 1 A typical drift mining process cycle

mechanisation and automation. Historical data over the period of 1 year from an underground mine in Sweden show that more than 15 percent of unplanned downtime of mobile equipment is related to the drilling rigs, with the greater part of the downtime attributed to the poor reliability of their components or subsystems. Other factors contributing to downtime include the harsh work environment and the operating context. Given this combination of factors, the drilling rig often represents a bottleneck in the mining production cycle and, thus, is becoming an important research topic.

3 Data Collection and Analysis

The failure data used in this paper were collected over a period of 2 years (2009–2011). The source of the data is the database of an underground mine in Sweden participating in the study. This database belongs to the MAXIMO system, a computerised maintenance management system (CMMS). In this research study, the time to failure data (TTF data) and the time to repair data (TTR data) of three drilling rigs and their subsystems were arranged in chronological order so that statistical analysis could find trends in the failure and repair data.

The first step in analysing the data was calculation of the times between failures (TBFs) for the system. In the CMMS, the failure data are recorded based on calendar time. Since drilling is not a continuous process, the TBFs were estimated by considering the utilisation of each rig. Reliability and maintainability data analysis is usually based on the assumption that the TBF and TTR data are independent and identically distributed (iid) in the time domain. It was critical to conduct a formal verification analysis of the assumption that the TBF and TTR data were iid; otherwise completely wrong conclusions could be drawn [13, 14]. Accordingly, the next step, after sorting and classifying the TBF and TTR data based on the subsystem level, was validation of the iid assumption. The failure data were tested for trends with the Laplace trend test. This test is used to determine whether a data set is identically distributed [14]. If such a trend is observed, classical statistical techniques for reliability analysis may not be appropriate, and a non-stationary model such as the non-homogenous Poisson process (NHPP) must be fitted [13–18]. Otherwise, the serial correlation test can be used to test the dependence of the failure data. A dependence test determines whether successive failures are dependent in data without a long-term trend [14]. If a dependence between successive failure data is observed, a branching Poisson process (BPP) model can be used [13]. If dependence is not observed, the iid assumption is valid. In this study, after testing the validity of the iid assumption, we examined different types of statistical distributions and estimated their parameters using the Easy Fit and Minitab software. The goodness of fit of the distribution was tested by using the Kolmogorov-Smirnov (K-S) test with the Easy Fit software. In the present paper, all statistical tests used a significance level (α) equal to 0.05.

4 Reliability Analysis Using Monte Carlo Simulation Method

Following the process described above, after classifying and sorting the data, we calculated the TBF of each subsystem and performed statistical validation to look for the presence of structures or trends in the failure data using Laplace trend and serial correlation tests. The Laplace trend test was used to test the hypothesis that a trend did not exist within the TBF data. We calculated the test statistic U for the TBFs of the drilling rig subsystems to be at a significant level of 0.05. From the standard normal tables, with a significant level of 0.05, the critical value is equal to 1.96. If $-1.96 < U < 1.96$, we accepted the hypothesis of no trend within the TBF data. After applying the trend test for the critical components of the studied rigs, we found no trend within the TBF data; for example, U was equal to 0.55 in the feeder of rig A used in the collaborating mine.

We performed a serial correlation test of the TBF data of the drilling rigs and their subsystems to check the dependence of the TBF data. To this end, we plotted the ith TBF against the (i-1)th TBF. We then tested the significance of the correlation by calculating the (r) value and comparing it with the critical (r) value obtained from the correlation tables. The results of the serial correlation test of the above component (i.e. the feeder) are given in Fig. 2. Since the points in the figure

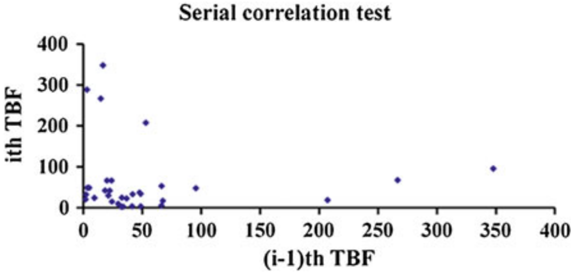


Fig. 2 Serial correlation test for the feeder of the drilling rig A

are randomly scattered, the failure data can be assumed to be independently distributed. The (r) value is equal to 0.05, while the critical value of (r) from the correlation tables at the significance level alpha = 0.05 and a degree of freedom of 30 for a two-tailed test is equal to 0.364. We can conclude that the correlation between the TBFs of this subsystem is statistically not significant (i.e. no correlation exists), since r < the critical r.

The results of the tests showed all subsystems in three studied rigs are free from trends and serial correlations and are identically and independently distributed (iid); therefore, the renewal process is the best way to perform reliability analysis on these subsystems. We examined various types of statistical distributions and estimated their parameters using the Easy Fit and Minitab software. The best fitted failure density functions are shown in Tables 1, 2 and 3.

The Monte Carlo simulation method plays an important role in system reliability assessment and optimal maintenance of large-scale complex networks but, in general, there are four major difficulties in evaluation [5]:

- System reliability structure may be very complicated;
- Subsystems may follow different failure distributions;
- Subsystems may have arbitrary failure and repair distributions for maintained systems; and
- Failure data of subsystems are sometimes not sufficient and sample size of life test or field population tends to be small.

Table 1 Data analysis of subsystems of rig A

Rig A		
Subsystems	Best fitted function	Parameters
Hoses	Weibull 2P	α = 0.92
		β = 20.75
Rock drill	Weibull 2P	α = 0.98
		β = 69.1
Feeder	Lognormal 3P	σ = 1.26
		μ = 3.4
		γ = −0.14
Boom	Weibull 2P	α = 1.04
		β = 146.28
Accumulators	Normal	σ = 197.41
		μ = 256.16
Hydraulic system	Gamma	α = 0,336
		β = 1047
Valves	Lognormal	σ = 1.17
		μ = 4.36
Control panel	Exponential	λ = 0.008
Water system	Lognormal	σ = 1.27
		μ = 5.17

Table 2 Data analysis of subsystems of rig B

Rig B		
Subsystems	Best fitted function	Parameters
Hoses	Weibull 3P	α = 0.95
		β = 55.53
		γ = 0.6
Rock drill	Lognormal	σ = 1.26
		μ = 3,27
Feeder	Weibull 2P	α = 0.82
		β = 42.47
Boom	Exponential	λ = 0.006
Accumulators	Normal	σ = 214.1
		μ = 300.5
Cable system	Weibull 2P	α = 1.09
		β = 339.7
Hydraulic system	Weibull 3P	α = 0.6
		β = 148.3
		γ = 16.92
Steering system	Weibull 3P	α = 1.15
		β = 112.9
		γ = 4.27

Table 3 Data analysis of subsystems of rig C

Rig C		
Subsystems	Best fitted function	Parameters
Hoses	Lognormal 3P	σ = 1.072
		μ = 3.12
		γ = −1.19
Rock drill	Gamma 3P	α = 1.13
		β = 52.61
Feeder	Exponential	λ = 0.018
Boom	Weibull 3P	α = 0.58
		β = 122.7
		γ = 19.04
Accumulators	Weibull 2P	α = 1.48
		β = 502.1
Cable system	Exponential	λ = 0.002
Hydraulic system	Lognormal 3P	σ = 0.77
		μ = 5.45
		γ = −66.72
Steering system	Lognormal 3P	σ = 0.62
		μ = 5.22
		γ = −37.7
Generator	Weibull 2P	α = 0.999
		β = 299.82

Among the various Monte Carlo reliability simulation algorithms, the K-R algorithm developed by Kamat and Riley [19] can be considered the most general and basic; other suggested methods for reliability simulation are merely modified forms of this method [20]. Therefore, the K-R method has been used for the reliability simulation of drilling rigs in this paper.

In this method, the failure times for individual components are generated based on the defined failure distribution function and then used to determine the success or failure of the system. The stages of the K-R method are [19]:

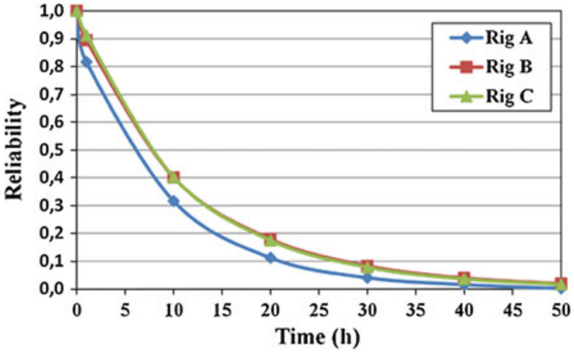
1. Find all minimal tie-sets from system reliability block diagram (RBD). Assume we must obtain system reliability interval estimates at some time point t.
2. From the life distribution of each subsystem, generate a random failure time t_i where i represents the ith subsystem, $0 < i <$ n.
3. Compare t_i with t for all subsystems. If $t_i >$ t, this indicates that at the time, t subsystem i functions properly; if $t_i < t$, the subsystem i has failed.
4. Determine whether the whole system is functioning or down according to the statues of its subsystems at t from step (3). Check all subsystems in a minimal tie-set. If all are operational, the system is operating properly at time t. If one or more fails, the tie-set is broken (failure) at t. Check the next minimal tie-set until an unbroken one appears, which means the system is operational at t. If all minimal tie-sets are broken, the system fails at t.
5. Repeat steps (2), (3), (4) for, say, N times. Count failure and success numbers of the system respectively: N_S (t) and N_F (t). Note that $N = N_S(t) + N_F(t)$.
6. The system reliability point estimate corresponding to t is given by Eq. (1):

$$\hat{R}(t) = \frac{N_S(t)}{N_S(t) + N_F(t)} \tag{1}$$

5 Results and Discussion

To ensure fast and reliable calculations during the simulation process, we prepared a computer program using MATLAB$^{\text{TM}}$ software. For each rig, we ran the program for different operation times with the iteration number of 10000 and achieved a reliability plot for each. Figure 3 shows the reliability plots of all three rigs achieved using the simulation method in one area. As can be seen in this figure, rig A has the lowest reliability of the three drilling rigs. However, the difference is small; the maximum value is 10 %, at about 15 h. All studied rigs are almost equal in reliability in the period of high reliability operation (from time 0 to 5 h) and in the

Fig. 3 Reliability plots of all rigs



period of very low reliability operation (after 35 h). The reliability of all rigs decreases by almost zero after 50 h. The main reason for this result is that the collaborating mining company bought the three rigs in the period 2003–2005 but kept failure and repair data in CMMS only from 2009. Therefore, the rigs were already in the wear-out failure period when the data were collected for this study; see Fig. 4. It is also obvious from Fig. 3 that the reliability plot of rigs B and C are so extremely close that they are almost the same.

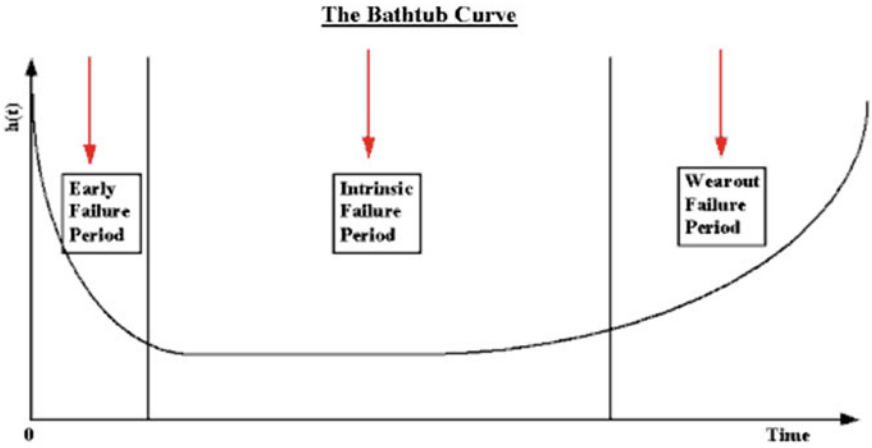


Fig. 4 Bathtub curve (adapted from [21])

6 Conclusions

In this paper, we analysed the reliability of three drilling rigs in a Swedish underground mine using the Kamat-Riley (K-R) simulation method. We ran the simulation process based on the series configuration of the repairable subsystems for the drilling rigs. To set up the simulation process, we created a computer program in MATLAB$^{\text{TM}}$ software. The results of simulation suggested the reliability of all rigs reduces to zero at about 50 h, possibly because the three rigs were already in their wear-out failure period when the data were collected for this study. As this short time shows, this important mining machine needs serious maintenance and servicing planning to reduce its downtime. Our overall aim was to test the applicability of the Monte Carlo simulation method to the analysis of the rig's reliability; we found this method is appropriate for reliability studies. It is time consuming, however, and hence best suited for large, complicated systems. Future studies should consider a comprehensive examination of maintenance scheduling and cost analysis of the underground drilling rigs.

Acknowledgments The maintenance personnel and maintenance management system employees at Boliden AB and Atlas Copco who helped the authors in this research are thanked for their support.

References

1. Al-Chalabi H, Lundberg J, Wijaya A, Ghodrati B (2014) Downtime analysis of drilling machines and suggestions for improvements. J Qual Maint Eng 20(4):306–332
2. Al-Chalabi H, Lundberg J, Jonsson A, Ahmadi A (2014) Case Study: Model for economic lifetime of drilling machines in the Swedish mining industry. Eng Econ. doi:10.1080/0013791X.2014.952466
3. Krishna MK, Verma AK, Srividya A, Ljubisa P (2010) Integration of black-box and white-box modeling approaches for software reliability estimation. Int J Reliab Qual Saf Eng 17(3):261–273
4. Hoseinie SH, Ataie M, Khalokakaie R, Ghodrati B, Kumar U (2012) Reliability analysis of cable system of drum shearer using power law process model. Int J Min Reclam Environ 26 (4):309–323
5. Wang H, Pham H (1997) Survey of reliability and availability evaluation of complex networks using monte carlo techniques. Microelectron Reliab 37(2):187–209
6. Kumar U, Klefsjö B, Granholm S (1989) Reliability investigation for a fleet of load haul dump machines in a Swedish mine. Reliability engineering system safety 26(4):341–361
7. Kumar D, Klefsjo B, Kumar U (1992) Reliability analysis of power transmission cables of electric mine loaders using the proportional hazards model. Reliab Eng Syst Saf 37(3):217–222
8. Vagenas N, Nuziale T (2001) Genetic algorithms for reliability assessment of mining equipment. J Qual Maint Eng 7(4):302–311
9. Vayenas N, Xiangxi W (2009) Maintenance and reliability analysis of a fleet of load-haul-dump vehicles in an underground hard rock mine. Int J Min Reclam Environ 23 (3):227–238

10. Gustafson A, Schunnesson H, Galar D, Kumar U (2012) The influence of the operating environment on manual and automated load-haul-dump machine: a fault tree analysis. Int J Min Reclam Environ. doi:10.1080/1755182X.2011.651371
11. Hoseinie SH, Ataei M, Khalokakaie R, Ghodrati B, Kumar U (2012) Reliability analysis of drum shearer machine at mechanized longwall mines. J Qual Maint Eng 18(1):98–119
12. Al-Chalabi H (2014) Reliability and life cycle cost modelling of mining drilling rigs. Doctoral thesis, Lulea University of Technology, Sweden
13. Ascher H, Feingold H (1984) Repairable systems reliability: modeling, inference, misconceptions and their causes. Marcell Dekker, New York
14. Kumar U, Klefsjö B (1992) Reliability analysis of hydraulic systems of LHD machines using the power law process model. Reliab Eng Syst Saf 35(3):217–224
15. Modarres M (2006) Risk analysis in engineering: techniques, tools, and trends. Taylor & Francis Group, Boca Raton
16. Birolini A (2007) Reliability engineering: theory and PRACTICE, 5th edn. Springer, Heidelberg
17. Louit DM, Pascual R, Jardine AK (2009) A practical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data. Reliab Eng Syst Saf 94(10):1618–1628
18. Ghosh S, Majumdar SK (2011) Reliability modeling and prediction using classical and Bayesian approach: a case study. Int J Qual Reliab Manage 28(5):556–586
19. Kamat SJ, Riley MW (1975) Determination of reliability using event-based Monte Carlo simulation. IEEE Trans Reliab 24(1):73–75
20. Hoseinie SH, Khalokakaie R, Ataie M, Ghodrati B, Kumar U (2013) Monte Carlo reliability simulation of coal shearer machine. Int J Perform Eng 9(5):487–494
21. Blischke W, Murthy D (2003) Case studies in reliability and maintenance. Wiley, Hoboken

Comparison of Mine Production Index Factors for Rock Bolter and Shovel

Amol Lanke and Behzad Ghodrati

Abstract MPi uses availability, utilisation and performance to compare and evaluate equipment. These measures apply to all equipment in mining. However equipment use in mining operations could lead to different evaluation of its availability, utilisation and production performance. MPi evaluation and comparison of equipment on MPi basis thus becomes difficult. Shovel and rock bolters are equipment used in open pit and underground mining respectively. Rock bolters are used to place bolts in mining room. Bolts reinforce rock masses through restraining the deformation within the rock masses. In open pit mining shovels are used for loading broken rock in truck for hauling. MPi can be used as scale for comparison of these equipment. However the operational difference between two equipment leads to different evaluation of MPi. Using hypothetical case study difference between availability, utilisation and performance of the rock bolters and shovel is studied. It was found that these parameters must be measured differently for rock bolters than shovel. Rock bolters availability and utilisation should be given less weights consideration as compared to other equipment in continuous mining operation. Its performance measurement it dependent upon bolts installed and capacity of bolts installation rather than tonnage of ore produced. This study can be helpful for evaluating MPi for equipment which are non-continuously used and lacks output in terms of ore tonnage.

Keywords Mine production index · Shovel · Rock bolter · Productivity in mining · Mpi evaluation

A. Lanke (✉) · B. Ghodrati
Luleå University of Technology, Luleå, Sweden
e-mail: amol.lanke@ltu.se

B. Ghodrati
e-mail: Behzad.ghodrati@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_47

1 Introduction

Probing, outlining and collecting information about ore body leads to further analysis for which mining method is suitable starts. Extraction of minerals carried out beneath the earth surface is termed as underground mining, whereas extraction of minerals by digging the earth is termed as open pit mining. Underground methods are employed when the depth of the deposit, the stripping ratio of overburden to ore (or coal or stone), or both become excessive for surface exploitation. Block of earth are extracted from surface to retrieve the ore contained within. During the extraction process surface of earth is being continuously excavated, thus forming deep pit. This method of mining is open pit mining. Considering such different approach to achieve the minerals, both of these methods requires usage of different types of equipment. In any mining operation the total output, which can be termed as productivity is based on the productivity of equipment used. Overall productivity measurement in mining in difficult. Various methods have been used for productivity measurement in mining industry; these include use of data envelopment analysis (DEA) process, use of automation, and improvement of existing equipment. [1–4]

The comparison of productivity by various equipment leads to know the bottleneck equipment in operation. However this comparison is complicated since the capacities and the fleet size of each of the equipment are different. It seems that there is lack of single methodology to compare productivity by all equipment together in mining industry. Since mining is continuous operation, the overall productivity of mining operation is based on each and every equipment used in system. In order to compare these different equipment and compare them on a single scale, Mine producton index (MPi) was introduced by Lanke et al. [5].

MPi is scale defined with modification of the Overall Equipment Effectiveness. Quality parameter in its definition does not apply to mining industry and the utilization is measure differently for mining equipment [6]. Considering the implication of the managerial decision (such as fleet size, working hours etc.) the weights are added to each parameter in OEE equation.

The MPi equation thus formed is given as follows.

$$MPi = Av^{a} \times Ut^{b} \times Pp^{c} \tag{1}$$

where

Av availability of equipment

Ut Utilisation of equipment

And Pp Production performance of the equipment

a,b,c are the weights assigned such that a + b + c = 1 and 0 < a,b,c <1

While measuring the availability in mining operations for MPi, the basic assumption is that MPi is calculated over the calendar period. That is during the total working time the machine is in possession however failures avoid the use of equipment for production.

The MPi is applicable for most of the equipment in underground as well as open pit mining, since these elements are common for the most of the equipment. However for some equipment in mining all elements of MPi equation may not be measureable or even available for measurement. Hence the applicability of MPi must be checked for such equipment. This study thus tries to compare two equipment, one from the underground mining (rock bolters) and one from the open pit mining (shovels). These comparisons will also represent what elements could be considered for modification of MPi for application in underground mining, specifically for equipment whose characteristics are limited. Problem statement for this study can be summarized as follows. Evaluation of MPi for different equipment is possible. However due to nature of operations performed by rock bolters factors involved in MPi equation may calculated differently. This study will show comparison of MPi factors for shovel and rock bolters. This will also help in knowing, for what type of instruments MPi evaluation should be done differently. This comparison will also form a guideline for MPi evaluation of these type of equipment.

To know the difference between why and how MPi factors (Availability, Utilisation and Performance) differ for rock bolters and shovels, it is essential to know the operation sequence where these both equipment are used. Shovels are used in open pit mines, in open pit mining after the blasting operation the ore generated needs to pick up and loaded for further processing. Shovel with high capacities are important equipment for loading blasted ore to trucks for further processing. During the continuous mining, fleet of shovels are used for achieving high output. After the initial operation of drilling and blasting, shovel continuously load the material in trucks, which is then further processed. It is thus beneficial to run the shovels continuously to produce high amount of ore.

Rock bolters are used in underground mining operations. Following operation sequence in sub level caving- an underground mining method shows where the rock bolts are used.

1. Preparation of areas for ore extraction is called development.
2. Building of tunnels or process of drifting, this process is essential for creating transport roads.
3. Production drilling, during which the drifts are drilled with various types of fans.
4. Blasting, once drifts are drilled, they are filled with explosives and are blasted. Wherever necessary after drifting and blasting process the ceiling and walls are reinforced with rock bolters.
5. The blasted ore is then loaded with help of bucket loaders and carried out.
6. The ore from loaders is then transported to crusher for further processing.

Another type of underground mining is long wall mining. Mechanized shearers are used to cut and remove the coal at the face of the mine. The coal is then carried to surface for further processing. Similarity bolting operation is performed inside the drifts to continue the operation safely without falling of ceiling.

The room and pillar mining is most common method of coal mining. The network of room is established by cutting coal seams. As the rooms are cut continuous loading of coal on shuttle car or ram is for further processing. “Pillars” composed of coal are left behind to support the roof of the mine. As mining continues, roof bolts are placed in the ceiling to avoid ceiling collapse.

Therefor rock bolters are used for stabilizing excavated rock mass and tunnelling. They usually symmetrically arranged for transferring the load from unstable surface or exterior of the rock to stronger part of the rock. The rock mass can be reinforced by different typed of rock bolts.

2 MPi Factors Consideration for Shovel and Rock Bolters

To compare the production for two different types of equipment, a base scale must be defined. For this purpose MPi is defined with three elements of availability, utilisation and performance. To calculate the MPi, for each equipment, the study tries to summarizes comparison between these elements for shovel and rock bolters. Mining operation sequence described shows the difference between the availability, utilisation and performance of these equipment. How it can be compared and considered for evaluation of MPi for each equipment is discussed as follows.

2.1 Availability Considerations

Operational availability is ratio of time of the equipment available for the operation to the total operational time [7]. It is calculated by following equation [8],

$$\text{Availability} = \left[\frac{TH - DT}{TH}\right] \tag{2}$$

where

TH Total Hours
DT Downtime Hours

Availability can be measured on scale of calendar time meaning availability based on the total working hours and the downtime hours. The total working hours are calendar hours for all equipment. Any downtime loss is due to failure of equipment or due to planned stoppage for maintenance of equipment.

In an open pit mining operations failure of shovel leads to stoppage of loading and hauling operation in mining. Unless the shovel failures are corrected production stoppage occurs due to shovels. Failure correction of shovel has to be done either online or thorough the process of overhauling schedule. Online failure means failure of equipment during the operation of equipment, whereas offline failure is attributed when equipment is not in use and subjected to failure. A single shovel can be subjected to overhaul process considering the number of shovels (fleet of shovels) available for production.

Table 1 Effect of downtime on shovels and rock bolters, availability during operation

Downtime cause	Shovels	Rock bolters
Sudden failure (online)	Loss of production/stoppage of operations	Loss of production/stoppage of operations
Sudden failure (offline)	No loss of production, availability is un affected	No loss of production, availability is not affected
Planned downtime	Reduction in fleet capacity	May not cause reduction in capacity

Rock bolters are used in mining operation on intermittent basis. Once the rock bolters are used for specific period of time it will be stored until required for further operation. However if rock bolters is not available i.e. subjected to failure before starting of loading, may cause delay of further operations. This will lead to loss of overall productivity.

Availability is affected by downtime of equipment, which in turns affects the overall operation. Effect of downtime on the availability and the operation of these equipment as is shown in Table 1

However if the planned downtime for rock bolters combined with the standby time when rock bolters are not in operation, the sudden failure for rock bolters can be reduced significantly. This will increase the availability. Shovel is subjected to continuous operation this might causes downtime of shovel will be higher than that of the rock bolters.

2.2 Utilisation Considerations

Utilisation of mining equipment can be defined as "percentage of total time that an asset is scheduled to operate during a given time period expressed as a percentage. The time period is generally taken to be the Total Available Time (i.e., one year)." [9]. Utilisation can be calculated by Eq. (3), [8]

$$\text{Utilisation} = \left[\frac{TH - DT - SH}{TH - DT}\right] \tag{3}$$

where
SH Standby Hours

Utilisation is thus based on the downtime hours and standby hours. The standby hours could be due to mine planning, location of equipment for required operation, legislative reasons (for example change of operator is enforced due to legislative reasons) and operational requirement.

In case of shovel standby time could be due to the following reasons:

1. Standby time due to operational constraints (non-availability of ore/truck/operator etc.)
2. Standby time due to legislative reasons (mandatory change of operators etc.)
3. Standby time due to non-usage of equipment

The standby time of shovel is mostly affected by the waiting on trucks [10]. To eliminate such standby time of shovel for continuous operations match factors is proposed. The term match factor is usually defined as the ratio of truck arrival rate to loader service time [11]. The variables to reduce the standby time for shovels are complicated due to nature of its operation.

In case of rock bolters standby hours can be broken in two parts:

1. Standby time due to operational constraints (rock hardness, failing of bolts, incorrect planning)
2. Standby time after the completion of operation. (non-usage time)

Rock bolters in practice do not need frequent operator change. However their utilisation might be hampered due to limited availability of working faces [12]. If the standby time during the non-usage of rock bolters is considered for calculation of utilisation, its utilisation could be affected in negative manner.

For any equipment in mining the utilisation can be increased by reducing the downtime by performing optimal maintenance during planned maintenance time. However for rock bolters the planned maintenance can be scheduled during the standby time when it is not in operation. This could help achieve higher utilisation of rock bolters during the operation. Variation of standby time for rock bolters is less complicated as it is based on less number of factors.

2.3 Performance Considerations

The performance of an equipment can be defined as "The ability of an item to meet a service demand of given quantitative characteristics, this performance is based on the capability and availability of an equipment " [13],

The production performance is given by Eq. (4), [8]

$$\text{Production Performance} = \frac{\left[\frac{AP}{TH-DT-SH}\right]}{RC} \tag{4}$$

where

AP Total actual output by equipment

RC Rated capacity of the equipment

For shovel the actual output and rated capacity can be given in terms of tonnage of ore. Production performance for shovel can be measured in actual output in tonnage when shovel utilisation is carried out to the overall capacity of the equipment.

However for rock bolters, there is no direct output in terms of ore tonnage. Although output by rock bolters help increase the tonnage of ore, it does not help is ore production directly.

Thus production performance of the rock bolter will depend upon the two factors [12]

1. Time for bolt installation (cycle time per bolt installation)
2. Installation capacity (Number of bolt installation possible in a given time period)

The actual output by rock bolters is number of bolts installed in a given period, and rated capacity can be total capacity over given period of time. Production performance equation for rock bolters (RB) thus can be given as the Eq. (5),

$$\text{Production performance} = \left[\frac{\frac{Actual\,bolt\,installations}{TH - DT - SH}}{Bolt\,instllations\,capcity\,spread\,over\,given\,time}\right] \tag{5}$$

During the calculation of performance for shovel and the rock bolters although the equation parameters have been changed the formula remains the same. Analysis of MPi evaluation for rock bolters and shovels

3 MPi Factors Evaluation for Shovel and Rock Bolter

In this section we will analyse the hypothetical case study. The following assumptions are made

1. There is only one shovel and one rock bolters, working in open pit and underground mines respectively.
2. The time considered is period of 1 day that is 24 total working hours.
3. The availability, utilisation and performance data for shovel is taken from an actual open pit mine database, however all data for rock bolters is assumed.
4. The weights for shovel are obtained through different case study for evaluation of its MPi.

On scale of calendar time rock bolters availability, utilisation, downtime and standby hours can be shown as in Table 2.

Table 2 Distribution of hours for shovels and rock bolters for hypothetical case study

Time	Shovel	Rock bolter
Availability	24	24
Downtime (during operation)	7.03	0.30
Downtime (after operation)	–	2.30
Utilisation hours	12.30	5
Idle time (during operation)	4.27	0
Idle time (non-operational)	–	16.00

The total hours of work are 24 h (divided in three shifts each of 8 h).Actual availability is attained when the downtime during and after the operation is considered. The availability calculation according to Eq. (2) is

$$\mathrm{AV_{SH}} = \left[\frac{24 - (7.03)}{24}\right] \times 100 = 69.04\,\%$$
$$\mathrm{AV_{RB}} = \left[\frac{24 - (0.30 + 2.30)}{24}\right] \times 100 = 83.33\,\%$$

According to utilisation definition, time it is schedule to complete the task is 5 h for rock bolters, whereas for shovels this is continuous operation during given 24 h, it is utilised unless there is standby or due to failure. For the rock bolters during its 5 h utilisation there is no downtime or idle time. For shovels there has been downtime hours of 7.03 and idle hours are 4.27 (4 h 27 min).

Utilisation for both these equipment calculated using Eq. (3) as,

$$\mathrm{UT_{SH}} = \left[\frac{24 - 7.03 - 4.27}{24 - 7.03}\right] \times 100 = 78.89\,\%$$
$$\mathrm{UT_{RB}} = \left[\frac{24 - (0.30 + 2.30) - 16}{24 - (0.30 + 2.30)}\right] \times 100 = 23.80\,\%$$

For measuring the performance of shovel the amount of ore loaded by shovel into trucks and its capacity to load actual tonnage of ore are required. The performance evaluation according to Eq. (4) for both equipment can be done as follows,

For shovel

AP Total ore loaded by shovel in its utilisation period

RC total capacity of shovel to load the ore in trucks in total available time

$$\mathrm{PP_{SH}} = \frac{6677.5}{193806} \times 100 = 34.50\,\%$$

Now in case of the rock bolters, actual output and capacity both should be in terms of bolt installation i.e.

AP bolts installed during utilisation time of 5 h, and
RC capacity of rock bolters to install bolts during the total available time

$$\mathrm{PP_{RB}} = \frac{720}{1130} \times 100 = 63.71\,\%$$

4 Discussion on Case Study

Out of total 24 h, shovel is available for 15 h and 57 min compared to 21 h for the rock bolters. The availability of both equipment can be increased by reducing the downtime. Shovel failures while in operation (online failure) and offline failures can be avoided by performing the perfect maintenance. However this would require assigning of planned maintenance time for shovel. This will reduce its availability since shovel operation is continuous.

In case of rock bolters it is possible to eliminate/reduce the downtime by performing planned maintenance before or after operation during its standby (non-operational) time. It means that although the planned maintenance can increase availability for both equipment, shovel availability is affected highly than rock bolters. In maintenance planning shovel has less time for maintenance due to it continuous operational requirement, whereas rock bolters can be maintained during its longer non-operational standby time. This is important factor while assigning weights to rock bolter availability, as compared to other equipment in underground mining.

As it can be seen that utilisation percentage of rock bolters is 23.80 % and utilisation of shovel is 79 %. From this it could be concluded that rock bolter is less utilised. However rock bolter was utilised for total operational time with no standby during its operation. The utilisation of rock bolter is 5 h spread over the period of total available time of 21 h. In shown case if non-operational standby time of rock bolter is not considered, its utilisation percentage reaches 100 % which is theoretical limit.

The shovel utilisation is 12 h and 30 min over the working period of 15 h and 57 min. There was standby time of 4 h and 27 min during operation of shovel. However its overall standby time is equal to its only operational standby time.

It is seen that rock bolters utilisation percentage thus is not comparable and usable for MPi evaluation. Utilisation hours over operational hours seem to be comparable scale.

Although performance measurement element of shovel and rock bolter are different, their performance measurement is comparable.

Comparing these two equipment it can be this said that,

- Rock bolters availability appear higher than shovel.
  - But Shovel is subjected to continuous operation.
  - Rock bolters is subjected to intermittent operation.

- Utilization of rock bolters in lower than shovel.
  - However Rock bolters are subjected to high standby hours due to its operation.

5 Conclusion

To compare all the mining equipment together MPi can be used as common factor. However for some equipment used in mining availability, utilisation and performance, which are core factors for MPi evaluation, may not be the same. To illustrate these differences and to evaluate MPi, even with these differences, the study compared two such equipment in open pit and underground mining.

Rock bolters are used in underground mining for purpose of placing bolts over the ceiling of rock mass to achieve stability of structure for further mining operation. Shovels are used in open pit mines for excavating the ore and waste for loading and further processing. Rock bolters are used intermittent throughout the operation, whereas shovels are and can be used continuously in mining operations. The shovel output can be directly measured in tonnage of rock moved, whereas rock bolters output is in terms of bolts installed.

While comparing, the following differences were noted between these two equipment. Availability hours are total hours for which the equipment can be used for operation. It is dependent upon downtime of equipment. The availability of mining equipment can be kept high by reducing the downtime. Due to nature of operation, compared to shovel, rock bolters standby hours are higher. The downtime of rock bolters can be reduced by performing the maintenance during the standby hours. Thus effective maintenance of rock bolters is possible during its non-operational time. Due to nature of operation shovel is subjected it has less standby time, but could be subjected to more operational downtime.

Utilisation of mining equipment is based on standby hours and downtime hours. The standby hours for rock bolters are parts of its availability. After the completion of operation rock bolters are stored, whereas for shovels the standby time is due to ineffective operations. Thus considering the standby hours for utilisation evaluation of rock bolters may decrease its overall MPi value and not considering the standby hours might lead to higher overall MPi value. The comparison of utilisation percentage thus becomes complicated.

Rock bolters performance although not directly related to tonnage of ore produced, is certainly calculable. For rock bolters it is advisable measure actual production and rated capacity in relation to the bolt installation.

It seems that MPi calculation must be done differently for rock bolters than shovels. Authors recommend the following considerations while calculating and evaluating the MPi for rock bolters.

- It is possible to avoid rock bolter downtime effectively, by performing its maintenance during its standby time.

  Weight evaluation for availability (a) should considers the occurrence of downtime occurs for rock bolters. (i.e. if downtime occurred during or after operation). Rock bolter availability should be normalized considering operational requirement and downtime as a function of operation
- The weights considered for utilisation (b) of rock bolters should be based upon the knowledge that its actual utilisation is spread over total available hours. When evaluating MPi for rock bolter, weights assignment for utilisation should be based on the utilisation hours over the total available hours for all equipment than only utilisation percentage of the rock bolter i.e. Utilization of rock bolters evaluation should be normalized considering its non-operational requirement hours
- While calculating the performance of rock bolters, actual production in terms of bolts installed and capacity in terms of ability to install bolts over period of time stretched for the time used must be considered.

Acknowledgments Authors would like to acknowledge CAMM project for financial support of this research work.

References

1. Kulshreshtha M, Parikh JK (2002) Study of efficiency and productivity growth in opencast and underground coal mining in India: a DEA analysis. Energy Econ 24(5):439–453
2. Rodríguez XA, Arias C (2008) The effects of resource depletion on coal mining productivity. Energy Econ 30(2):397–408
3. Tsolas IE (2011) Performance assessment of mining operations using nonparametric production analysis: a bootstrapping approach in DEA. Res Policy 36(2):159–167
4. Brown GM, Elbacher BJ, Koellner WG (2000) Increased productivity with AC drives for mining excavators and haul trucks. Industry applications conference, 2000. conference record of the 2000 IEEE, vol 1. IEEE
5. Lanke A, Hoseinie H, Ghodrati B (2014) Mine production index (MPI): new method to evaluate effectiveness of mining machinery. International conference on mining and mineral engineering (ICMME 2014)
6. Paraszczak J (2005) Understanding and assessment of mining equipment effectiveness. Mining Technol 114(3):147–151
7. Macheret Y, Koehn P, Sparrow D (2005) Improving reliability and operational availability of military systems. Aerospace Conference, 2005 IEEE. IEEE
8. Dhillon BS (2008) Mining equipment reliability, maintainability, and safety. Springer Science & Business Media
9. The society for maintenance and reliability professionals. Metric definitions. http://library.smrp.org. Accessed 30th Jan 2015
10. Krzyzanowska J (2007) The impact of mixed fleet hauling on mining operations at Venetia mine. J S Afr Inst Min Metall 107:215–224
11. Burt CN, Caccetta L (2007) Match factor for heterogeneous truck and loader fleets. Int J Mining, Reclam Environ 21(4):262–270

12. Gustafson A, Schunnesson H, Ghosh R (2014) Bolting procedures in Outokumpu's Kemi mine. Mine Planning and Equipment Selection. Springer International Publishing, pp 411–420
13. IEC (1990) 60050 (191): dependability and quality of service. International Electrotechnical Commission, Geneva, Switzerland.

Reliability Analysis of Face-to-Surface Continuous Coal Hauling System in Longwall Mines

Amid Morshedlou and Hesam Dehghani

Abstract In this paper the reliability of the haulage system in Tabas coal mine has been discussed using the failure and failure interval data in past 2 years for Armored Face Conveyor (AFC), Beam Stage Loader (BSL) and conveyer belt. In respect to the data study and classification, conveyor belt with failure abundance of 50.5 % is the most critical, while AFC with the failure abundance of 22.3 % shows the best performance. The results of data analysis indicate that all three machines' reliability distribution function obeys the power law process. The reliability of AFC, BSL and conveyer belt reaches zero after 220, 30 and 8 h of continuous operation, respectively. The conveyor belt is the first system, which its reliability reaches to zero and cause the entire hauling operation to stop. Regarding to the high potential of failure in conveyer belt, it has always been considered as the most significant part of the system's failure hence should be monitored more precisely. So, the conveyer belt is the most critical subsystem of the haulage system. Approximately, the reliability of the haulage system after 4 h reaches nearly zero. In the first hour system's performance, it almost looses 90 % of its reliability, which is a considerable amount.

Keywords Reliability · Armoured face conveyor · Beam stage loader · Conveyor belt

1 Introduction

The importance of fossil fuels is increasing day by day due to their limited resources, therefore a great attention has been drawn to the industrial equipment related to them. Coal is one of the most important fossil fuels, which has many

A. Morshedlou · H. Dehghani (✉)
Hamedan University of Technology, Hamedan, Iran
e-mail: dehghani@hut.ac.ir

A. Morshedlou
e-mail: morshedlou@stu.hut.ac.ir

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_48

applications in the steel industry and power generation. Nowadays most of the world's coal is mined by mechanized long-wall mining method. The most important equipment used in these mines are drum shearer, Armoured Face Conveyor (AFC), Beam Stage Loader (BSL), Powered Supports and Conveyor Belt. Figure 1 shows a mechanized long-wall mine. After the coal is cut by drum shearer, materials load on AFC to be delivered to BSL. Eventually BSL delivers the coal to conveyor belt to transport it out from the stope. AFC, BSL and conveyor belt play the most significant roles in hauling operation and production process in long-wall mines. Therefore, their reliability is significant to keep mine production at the desired level and for maintaining smooth operation as well as achieving better production conditions.

Many researches have been conducted on reliability and maintenance of mining equipment. The application of reliability engineering in mining industries has been conducted since 1960. The initial studies have mostly used the qualitative approach and they only consist of descriptions about the machine failures and production delays. Nevertheless, mathematical and quantitative analysis methods have been used since the end of the 1980s. With the developments in new mining equipment, reliability analysis also became more complicated. Because of the two above-mentioned reasons, more reliability studies are required on the mining equipment. The reliability studies on longwall mining equipment during last two decades are being briefly reviewed below.

Mandal and Banik(1996) did a study on long-wall equipment in few Indian coalmines. They considered the AFC, shearer, stage loader and belt system as the components of production process. They calculated the delay hours, reliability, product loss, and presented the production failure risk for each subsystem [1]. Gupta et al. developed a method using maintenance information, such as, Time To

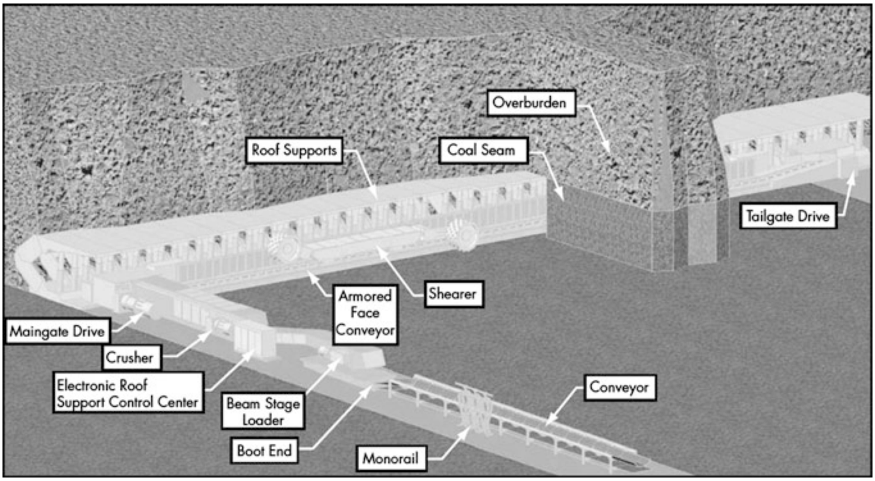


Fig. 1 Locating and components of long-wall mines

Failure (TTF) for pinpointing the weak links in the shearer machine. In this study a Fault Tree (FT) technique was used to understand the failure logic of a long-wall shearer and its components were ranked by the Birnbaum factor. The subsequent analysis showed that the results can be useful in applying a replacement policy as well as maintenance [2]. Gupta and Bhattacharya used the fault tree technique to study the reliability of AFC and presented the reliability curve. Main purpose of the paper was to explore the major weakness point of AFC and provide a good solution to reduce downtime problems [3]. Bing-Yuan et al. studied the reliability of the production system of long-wall face. They came up with theories to improve productivity, discussing the application of the transformation plan and optimization of a reasonable coal stock capacity as well as a selection of system equipment and matching optimization [4]. Hoseini et al. studied the reliability of water system of drum shearer with considering three subsystems in series network; filters, spray jets and hoses and valves. The result showed that the filters subsystem has the highest reliability importance among all, therefore is defined as the most critical subsystem [5]. Reliability and maintainability of electrical system of drum shearer was analyzed by Hoseini et al. The reliability-based maintenance intervals for 90, 80, 70 and 50 % reliability level were calculated. The calculations shows that the time to repair (TTR) of this system varies in range between 0.17 and 4 h [6]. Hoseini et al. developed a reliability model of hydraulic system of drum shearer. The model showed that the reliability of the hydraulic system reduces to zero value after approximately 1650 h of operation. The failure rate of this system decreases by the increase in time. Therefore, corrective maintenance (run-to-failure) was selected as the best maintenance strategy for it [7]. Reliability-based maintenance scheduling of haulage system of drum shearer has been modeled by Hoseini et al. The result showed that Time Between Failures (TBF) data of this system obeys the three-parameter Weibull distribution. Based on the achieved reliability model, the Preventive Maintenance (PM) scheduling has been suggested for different reliability levels [8]. Hoseini et al. used power law process to analyzed the reliability of cable system of drum shearer. Based on analysis and results, a period of 125 h was defined as the reliability-based maintenance interval for the cable system [9]. The reliability of drum shearer of Tabas coal mine was studied by Hoseini et al. with considering six subsystems in series network; Water system, haulage, electrical system, hydraulic system, cutting arms, and cable system. Pareto analysis showed that the water system is the most critical subsystem of the drum shearer. The failure rate analysis shows that the failure rates of the hydraulic, haulage and electrical systems decrease while the failure rates of the water system, cutting arms and cable system increase [10]. In further research Hoseini et al. used Monte-Carlo simulation for reliability analysis of water system of drum shearer and drum shearer itself with considering three and six subsystems, respectively [11, 12]. Morshedlou et al. studied the reliability of the AFC of the Tabas coal mine. They considered four subsystems in series network; electrical system, mechanical system, chain tension unite, speed control unite. Pareto analysis showed that the electrical system is the most critical subsystem of the AFC. Among the subsystems of the AFC, the speed control unite seems to be the most reliable subsystem [13]. The reliability of the

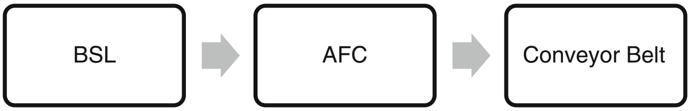


Fig. 2 Block diagram of hauling operation

electrical and mechanical units of the Tabas coal mine equipment was analyzed by Morshedlou et al. The results showed that the electrical units have a higher failure frequency and the reliability of the electrical units in all three equipment reduces to zero more quickly in comparison with mechanical units [14].

Regarding to the literature review, most of reliability studies have been done on underground mining machinery and systems. The complexity of machines, their large operational loads, and the harsh underground working conditions impose stringent requirements on reliability analysis for these machines. Very few of these studies talk about the subsystems of the long-wall equipment.

Hauling operation plays a significant role in the production process of long-wall mines. Other types of mining equipment and processes have been studied more correctly and their reliability has been discussed properly but there are neither sufficient nor applicable reliability studies on hauling operation for maintenance and operation management. Therefore, a fundamental study based on reliability characteristics is essential for improving the production and operation characteristics of hauling operation and whole long-wall system. Note that if a failure occurs in any of the machines, the entire hauling operation will stop, therefore their function is considered as series configuration. The block diagram of a hauling operation is shown in Fig. 2.

2 Reliability Analysis Process

The quantitative reliability analysis techniques use real failure data (obtained, for instance, from a test program or from field operations) in conjunction with suitable mathematical models to produce estimation of product or system reliability. Three stochastic processes are generally used for reliability analysis of repairable systems [9]:

(1) homogeneous Poisson process (HPP);
(2) renewal process (RP); and
(3) non homogeneous Poisson process (NHPP).

To determine which process is the best analysis method for available data, two analytical test were performed on the data. The first step is to perform a trend analysis to determine whether the data are identically distributed or not. Regarding to results of the trend analysis, if the assumption that the data is identically distributed is not valid, then classical statistical techniques for reliability analysis may not be appropriate; therefore, a non-stationary model such as non-homogeneous Poisson process (NHPP) must be fitted. The second step is to perform the correlation test on the data. If there is no trend and no serial correlation in failure data then the data is independent and identically distributed (iid). Classical statistical techniques are the best way for reliability modeling for iid data. The trend test can be made both analytically and graphically [1, 15]. There are five analytical methods for testing the presence of trend; Reverse Arrangement Test, Military Handbook Test, Laplace Test, likelihood-ratio test and Area Test. Military Handbook Test as one of the applicable analytic tests is better method at finding significance when the choice is between no trend and a NHPP Power Law model. This test checks the trend presence by calculating the test statistic *U* (Eq. 1) [16]:

$$U = 2\sum_{i=1}^{n} \mathrm{Ln}(T_n/T_i) \tag{1}$$

Where, n is total number of failures, T_n is time of the nth failure and Ti time of the ith failure. Under the null hypothesis of a HPP, the test statistic U is chi-squared distributed with $2(n - 1)$ degrees of freedom. If the null hypothesis be rejected at 5 % level of significance it means that the Time Between Failures (TBFs) data has trend and therefore, is not identically distributed [1].

In graphical methods, the trend test involves plotting the cumulative failure numbers against the cumulative time to failure. If the plotted points lie (or approximately) on a straight line, then the data is trend free and identically distributed (id). A test for serial correlation was also done by plotting the ith TBF against the (i−1)th TBF, i = 1, 2, …, n. If the plotted points are randomly scattered without any pattern, it can be interpreted that there is no correlation in general among the TBFs data and the data is independent.

The Kolmogorov-Smirnov (K-S) test is classically used for the validation and selection of the best-fit distribution [17]. The failure data analysis process, which is used in this study for selecting the best reliability based maintenance modeling, is shown in Fig. 3. Further explanations will be presented in the case study part.

3 Case Study

Tabas coal mine is located in the central desert of Iran and it is the largest long-wall coal mine of Iran. Parvadeh region covers an area of about 1200 km^2 located 70 km south of the city of Tabas. The region’s has largest coal reserves in country and its coal reserves estimated about 1.1 billion tones. The most suitable seam for mining (called C1) has 1.8 m thicknesses and is extracted by retreat long-wall method using a double-drum shearer. According to the production planning, mining production

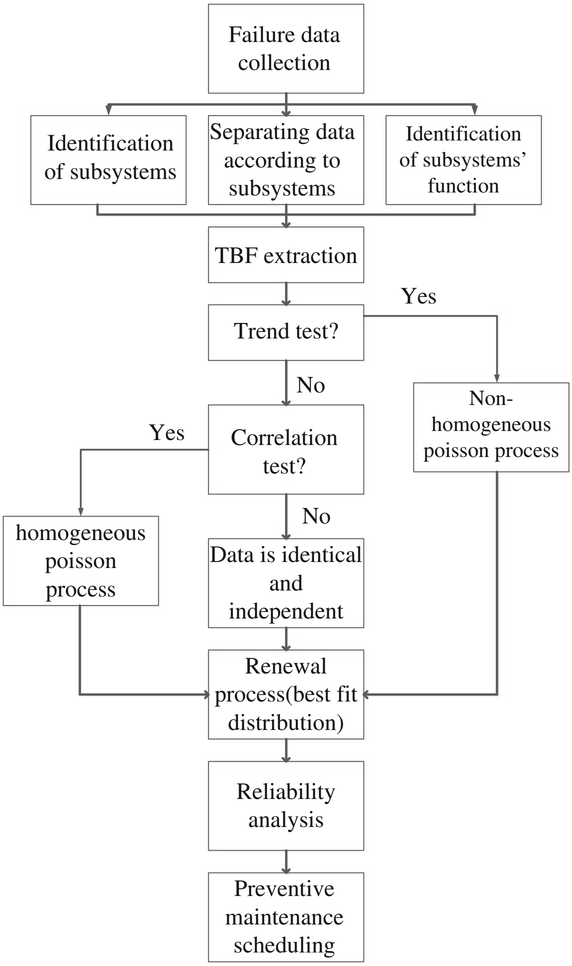

Failure data collection
Identification of subsystems
Separating data according to subsystems
Identification of subsystems' function
TBF extraction
Trend test?
Yes
Non-homogeneous poisson process
No
Correlation test?
Yes
homogeneous poisson process
No
Data is identical and independent
Renewal process(best fit distribution)
Reliability analysis
Preventive maintenance scheduling

Fig. 3 Reliability analysis process

rate should be 1.5 million tons per year at this time. As a result of the equipment failure Tabas coal mine had serious production problems, so the production process was experiencing consecutive downtimes. Due to this problem, the production rate reduced to approximately 500,000 tons per year. For example, in some shifts (8 h) the equipment would fail more than 10 times which would cause the production system stop for about 6 h. This means that the useful time of the production process was extremely low. The length of studied long-wall face is 215 m and panel length is 1200 m. Technical characteristics of AFC, BSL and conveyor belt of Tabas coal mine is presented in Tables 1, 2 and 3, respectively.

Table 1 Technical characteristics of AFC of Tabas coal mine

Parameters	Quantity
Conveyor length	219 m
Maximum capacity of conveyor	1300 t/h
Average capacity of conveyor	1000 t/h
Power of discharge drive	1 × 105/315 kW
Power of low return-end drive	1 × 105/315 kW
Speed of scraper chain	1.3 m/s
Pan height	295 mm
Pan width	842 mm
Pan length	1500 mm

Table 2 Technical characteristics of BSL of Tabas coal mine

Parameters	Quantity
Conveyor length	30.5 m
Maximum capacity of conveyor	1500 t/h
Average capacity of conveyor	1300 t/h
Power of discharge drive	1 × 55/160 kW
Speed of scraper chain	1.51 m/s
Pan height	260 mm
Pan width	846 mm
Pan length	1500 mm

Table 3 Technical characteristics of main conveyor belt of Tabas coal mine

Parameters	Quantity
Total power	2 × 224 kW
Belt width	150 cm
Belt speed	3 m/s
Capacity	800 t/h
length Belt	1250 m
Belt type	FR6000

4 Data Analysis

For identification of critical subsystem, Pareto analysis (failure frequency analysis) was done on the available data [18]. Figure 4 shows the results of Pareto analysis.

As seen in Fig. 4, conveyor belt have a higher failure frequency than the other equipment and consists of 50.2 % of all failures. This indicates that the most of the failures and production process stops in mine occurred due to the failure of the conveyor belt. In fact, the conveyor belt is the critical subsystem in terms of management issues related to maintenance and should be watched more carefully.

After data collection, the validation of the iid nature of the TBF data was performed. First, military handbook analytic trend test were applied on the data.

Fig. 4 Pareto analysis of hauling operation subsystems in Tabas coal mine

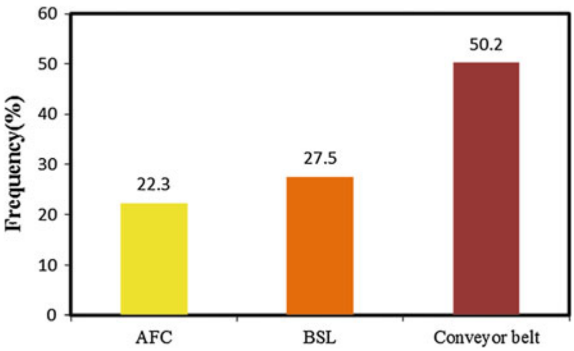


Table 4 The results of analytic test on subsystems of hauling operation

Equipment	Degree of freedom	Calculated statistic U	Lower chi2 value (2.5 % level of significance)	Upper chi2 value (97.5 % level of significance)
AFC	250	195.4	208.09	295.68
BSL	308	457.1	261.28	358.52
Conveyor belt	564	838.4	500.09	631.70

The computed values of the statistical test for the equipment are given in Table 4. Regarding to the results of analytical test on subsystems of hauling operation, the assumption that the failure data of subsystems does not follow any trend, is rejected for all machines. Consequently, the reliability of these subsystems should be analyzed by non-homogeneous Poisson process. In this study, power law process (PLP) is used for reliability modeling of all three machines.

5 Reliability Analysis

In order to calculate the best-fit distribution curve, Easyfit software was used. The Kolmogorov-Smirnov (K-S) test was used for selecting the best distribution among the top choices. The result of data analysis and best-fit distributions are illustrated in Table 5.

The reliability curves of AFC, BSL and conveyor belt were plotted using the, above-mentioned, distributions and their parameters, as illustrated in Figs. 5, 6 and 7, respectively. For further comparison and determination of the critical subsystems of the hauling operation in Tabas coal mine, the reliability curves of these machines are shown in Fig. 8.

As it can be seen in these figures, the reliability of AFC, BSL and conveyor belt reaches zero after about 220, 30 and 8 h of operation, respectively. Among the

Table 5 The results of data analysis and best-fit distributions

Equipment	Best fit distribution	Parameters	
AFC	Power law process	α = 1.289	β = 59.99
BSL	Power law process	α = 0.679	β = 2.645
Conveyor belt	Power law process	α = 0.675	β = 0.597

Fig. 5 The reliability plot of AFC

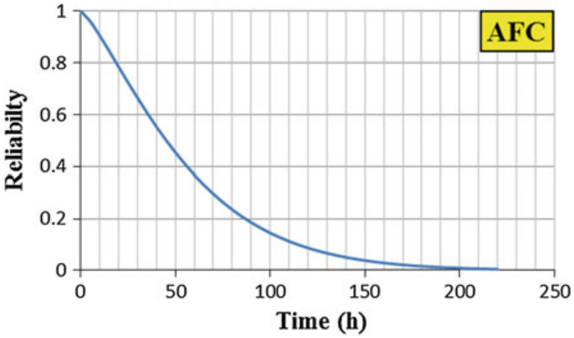


Fig. 6 The reliability plot of BSL

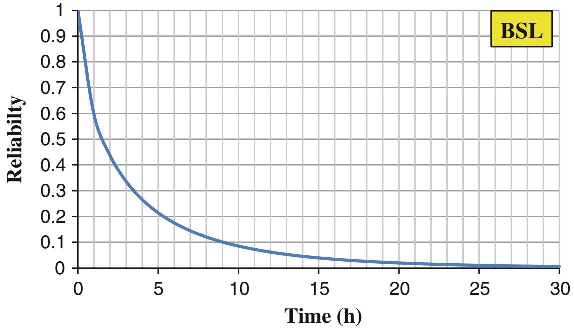


Fig. 7 The reliability plot of conveyor belt

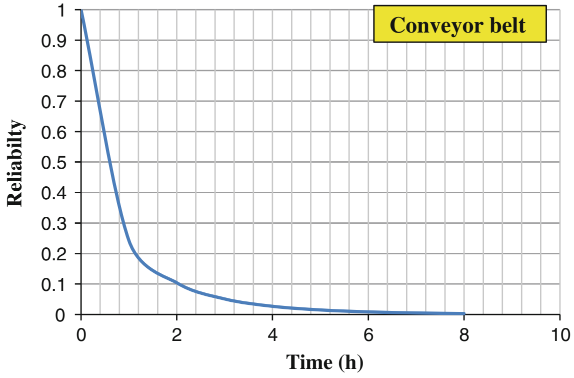
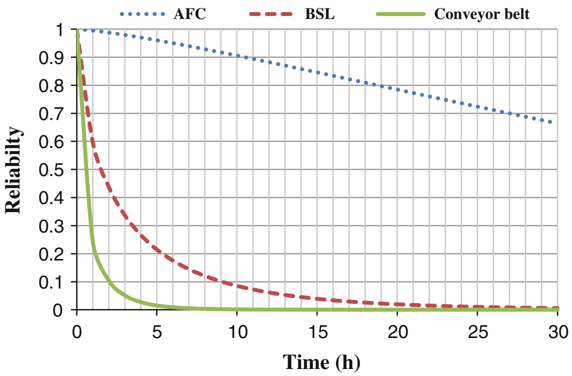


Fig. 8 The reliability plots of each subsystems of hauling operation



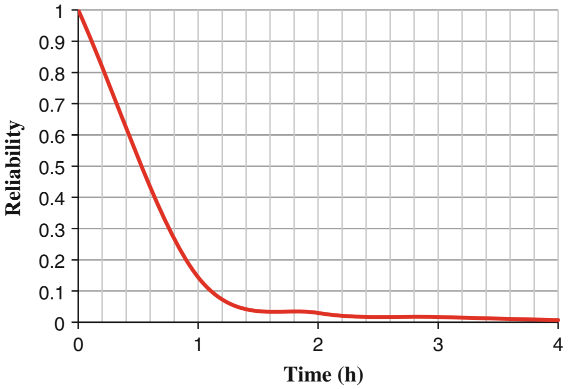
subsystems of the hauling operation, the AFC seems to be the most reliable subsystem and it takes about 45 h to reach the 50 % level of its reliability while the conveyor belt is the most critical subsystem of hauling operation and its reliability reaches zero before any other subsystems and it has the lowest reliability level in the machine operation.

Regarding the series configuration, the reliability of the hauling operation was calculated using a multiplier of the reliability of subsystems (Eq. 2):

$$\mathrm{R_{system}(t)} = \prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{R(t)}i \tag{2}$$

where the n is the number of subsystems and R(t)i shows the reliability of ith subsystem. Figure 9 shows the final reliability plot of hauling operation. As it can be seen in this figure, the reliability of the hauling operation reduces to zero in a period of about 4 h. There is a 50 % chance that the hauling operation will not fail for the first 0.5 h of operation. It shows that this operation needs serious attention and has high potential for causing the production stoppages, which is, the worst and the most critical threat for production continuity in long-wall mines.

Fig. 9 The reliability plot of hauling operation



6 Conclusion

Due to significant roles of hauling operation in continuity of production and extraction process in long-wall mines, assessing the reliability of AFC, BSL and conveyor belt is essential. In this paper Reliability of hauling operation of Tabas coal mine were investigated and the following results were obtained.

- The results of the Pareto analysis showed that the conveyor belt has the highest failure frequency among all the equipment and because of that it is the critical subsystem of hauling operation.
- The data analysis showed that the failure data in all three AFC, BSL and conveyor belt follow the assumption of non homogeneous Poisson process (power law process).
- The reliability of AFC, BSL and conveyor belt reaches zero after about 220, 30 and 8 h of operation, respectively.
- Among the subsystems of the hauling operation, the AFC seems to be the most reliable subsystem and the reliability of the conveyor belt reduces to zero more quickly than the other machines. Therefore it is the major factor to reduce the reliability of the hauling operation.
- The reliability of the hauling operation reduces to zero in a period of about 4 h. It shows that this operation needs serious attention and has high potential for causing the production stoppages, which is, the worst and the most critical threat for production continuity in long-wall mines.

References

1. Mandal SK, Banik PK (1996) Evaluation of reliability index of long-wall equipment systems. Min Technol 78(897):138–140
2. Gupta S, Ramkrishna N, Bhattacharya J (2006) Replacement and maintenance analysis of longwall shearer using fault tree technique. Min Technol 115(2):49–58
3. Gupta S, Bhattacharya J (2007) Reliability analysis of a conveyor system using hybrid data. Qual Reliab Eng Int 23:867–882
4. Bing-Yuan H, Gang S, Li-Xun K (2009) Reliability emulation of production system on long-wall face. J Coal Sci Eng 15(1):76–80
5. Hoseinie SH, Ataie M, Khalookakaei R, Kumar U (2011) Reliability modeling of water system of long-wall shearer machines. Arch Min Sci 56(2):291–302
6. Hoseinie SH, Ataie M, Khalookakaei R, Kumar U (2011) Reliability and maintainability analysis of electrical system of drum shearers. J Coal Sci Eng 17(2):192–197
7. Hoseinie SH, Ataie M, Khalookakaei R, Kumar U (2011) Reliability modeling of hydraulic system of drum shearer machine. J Coal Sci Eng 17(4):450–456
8. Hoseinie SH, Ataie M, Khalookakaei R, Kumar U (2011) Reliability-based maintenance scheduling of haulage system of shearer. Min Miner Eng 3(1):26–37
9. Hoseinie SH, Ataie M, Khalookakaei R, Ghodrati B, Kumar U (2012) Reliability analysis of the cable system of drum shearer using the power law process model. Int J Min Reclam Environ 1:1–15

10. Hoseinie SH, Ataie M, Khalookakaei R, Kumar U, Ghodrati B (2012) Reliability analysis of drum shearer machine at mechanized longwallmines. J Qual Maint Eng 18(1):98–119
11. Hoseinie SH, Ghodrati B, Kumar U (2013) Monte Carlo reliability simulation of water system of longwall shearer machine. Int J Reliab, Qual Saf Eng 20(6):1–11
12. Hoseinie SH, Ataie M, Khalookakaei R, Kumar U, Ghodrati B (2013) Monte Carlo reliability simulation of coal shearer machine. Int J Perform Eng 9(5):487–494
13. Morshedlou A, Dehghani H, Hoseini SH (2014) Reliability analysis of armoured face conveyor (AFC) in Tabas mechanized coal mine. 3rd international conference on reliability engineering, Tehran
14. Morshedlou A, Dehghani H, Hoseini SH (2014) Reliability-based maintenance scheduling of powered supports in Tabas mechanized coal mine. J Min Environ 5(2):113–120
15. Kumar U (1990) Reliability analysis of load-haul-dump machines. PhD thesis, Lulea University of Technology, Lulea
16. MIL-STD-2173 (1986) Reliability centered maintenance, Department of Defense, Washington
17. Kumar U, Klefsjo B (1992) Reliability analysis of hydraulic system of LHD machines using the power law process model. Reliab Eng Sys Saf 35(3):217–224
18. Barabady J, Kumar U (2007) Reliability analysis of mining equipment: a case study of a crushing plant at Jajarm bauxite mine in Iran. Reliab Eng Saf 93:647–653

Simulation of an Active Maintenance Policy: A Preliminary Study in Dragline Maintenance Optimization

Onur Gölbaşı and Nuray Demirel

Abstract Current maintenance policies for draglines do not cover enough preventive measures. Preventive maintenance for these systems is generally implemented via inspections and corrective activities unfortunately keep their priorities in dragline maintenance. Moreover, optimalities of both inspection intervals and their implementation durations are generally underestimated and they are determined via rough estimations. However, sustainability of a dragline operation and health of system components can be improved through maintenance optimization studies including preventive activities. This type of studies requires development of representative and comparable models to measure the effectiveness of optimization. In this sense, this paper presents the simulation of current maintenance policy of the draglines in Tunçbilek coal mine in Turkey, as a preliminary stage of the maintenance optimization study. The established policy aims to reveal 1-year halt profiles of the draglines via combining deterministic halts in operations and random lifetime characteristics of the system components.

Keywords Dragline · Lifetime characterization · Maintenance policy · Simulation

1 Introduction

Inherent risks in mining areas lead to unexpected and frequent failures of the mining equipments. Maintenance of these heavy-duty machines is performed typically via corrective repairing or replacement of the system components in malfunctioning state. Concern of recovering systems preventively and adaptation of

O. Gölbaşı (✉) · N. Demirel
Department of Mining Engineering, Middle East Technical University,
Ankara, Turkey
e-mail: golbasi@metu.edu.tr

N. Demirel
e-mail: ndemirel@metu.edu.tr

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_49

preventive measures to maintenance policies are still out of the desired levels. Moreover, inspections of machineries are generally carried out in non-optimal intervals without validating cost-effectiveness of these intervals. In mine sites, failures during lifetime of machinery systems are recorded roughly and these statistics are not benefitted sufficiently. However, these values offer a good interface in estimation of failure profiles of machines and improvement of effective maintenance policies. Analyzing failure data allows investigation of maintenance optimality via (i) estimation of root-causes of failures, (ii) assessment of system reliabilities, (iii) determination of preventive replacement decisions, (iv) optimization of overhauling or inspection intervals, and (v) specification of components that can be maintained simultaneously in opportunistic maintenance concept.

Maintenance policies cover various decisions on corrective and preventive activities to be applied along the lifetimes of systems. Optimization of these policies allows decision-maker to diminish overall operational expenses and to protect functional health of machinery components. A maintenance model can be built mathematically to maximize or minimize a prescribed objective function. In this respect, scope of the policy may aim to maximize performance factors such as reliability, availability, or profit or minimize other factors such as downtime, cost, or machinery deterioration. Moreover, these models may also consider both economic and downtime factors together to minimize unit lifetime cost of machinery while keeping availability above the limit values.

Draglines are one of the most complex systems employed in surface mines. They perform a single-handed overburden removal in open-cast mines and they are utilized alternative to truck and shovel system. Draglines are controlled using independent mechanisms of swing, hoist, and drag to excavate soft rock or loosen material after blasting via its bucket and to dump it onto an adjacent spoil pile [1]. These earthmovers are extensively utilized around the world. In the USA alone, 101 numbers of draglines with bucket capacity between 30 and 108 m^3 are employed in 56 surface coal mines and 40 % of overall overburden removal in open-cast mines is achieved by these machines [2]. Operational view of a dragline and its major subsystems can be viewed in Fig. 1.

Due to massive structure of dragline and rough ground conditions in excavation area, working components of the mechanism are frequently exposed to failures due to wear and tear, fractures, and fatigue. Any functionality loss in the components causes halt of dragline and delays in overburden stripping operations. Therefore, it is a requirement to constitute more conservative maintenance policies for these earthmovers to ensure continuity of operations and longevity of working components. This paper presents the preliminary stage of a dragline maintenance optimization study. In this sense, current maintenance policies applied for two draglines operating in Tunçbilek coal mine, Turkey, were simulated to achieve their annual failure profiles.

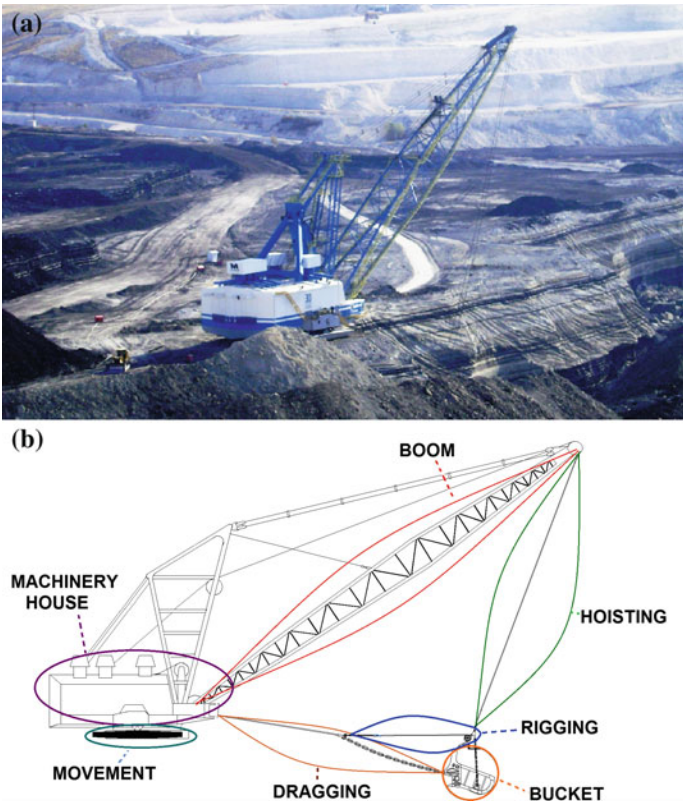


Fig. 1 Operational view (**a**) and main subsystems (**b**) of a dragline

2 Lifetime Characteristics of Dragline Components

Prior to maintenance simulation, it is required to estimate lifetime characteristics of dragline components to understand random failure behaviour in the system. In this sense, dragline was decomposed into seven main subsystems as given in Fig. 1. Then, major components inducing breakdowns were distributed to the subsystems, considering their failure modes and occurrence areas in the mechanism (Table 1). Following component distribution, relevant failure records were assigned to each individual component.

Datasets of components were initially analyzed for outlier detection using box plots. These plots essentially utilize six descriptive values as 1st quartile (Q_1), 2nd quartile (Q_2), 3th quartile (Q_3), maximum and minimum of the dataset, and number of observations. These quartiles indicate 25, 50, and 75th percentile points in data frequency curve, respectively. Outliers generally stay out of the range between $Q_1-1.5xIQR$ and $Q_3 + 1.5xIQR$ where IQR is the difference between 1st and 3th

Table 1 Major failure-inducing components in a dragline

Subsystems	Components
Dragging	Chain assembly, ringbolts, dragging rope, control, socket
Hoisting	Brake, hoisting rope, sockets, control
Bucket	Bucket body, chain assembly, digging teeth, pins, ringbolts
Rigging	Sockets, ringbolts, rigging rope, pulley
Machinery house	Generators, motors, lubrication system, air conditioning
Movement	Rotation mechanism, walking mechanism, warning mechanism
Boom	Boom chords

quartiles. Although boxplots offers a nonparametric test, independent of distribution type, this test cannot be used for highly-tailed, i.e. right skewed, failure frequency curves. Therefore, outliers in the study were generally detected subjectively according to general behavior of time-between-failures data, as well as boxplots.

After outlier elimination, each dataset of the working component was tested to check whether data is distributed identically and independently. In this sense, run charts offer an effective way to investigate data randomness via analyzing data anomalies such as clustering, mixture, trend, and oscillation (Fig. 2). Existence of any anomalies may point to undesired correlations between data values. It was observed from the run charts that there is not any potential threat against data randomness for dragline components except for data trend. In the study, trend behaviours of the datasets were also analyzed using hypothesis testing methods.

Lifetime trend behavior extensively effects the reliability assessment technique. Ascending or descending behavior of time-between-failures data refers an improvement or deterioration in the working mechanism of system, respectively.

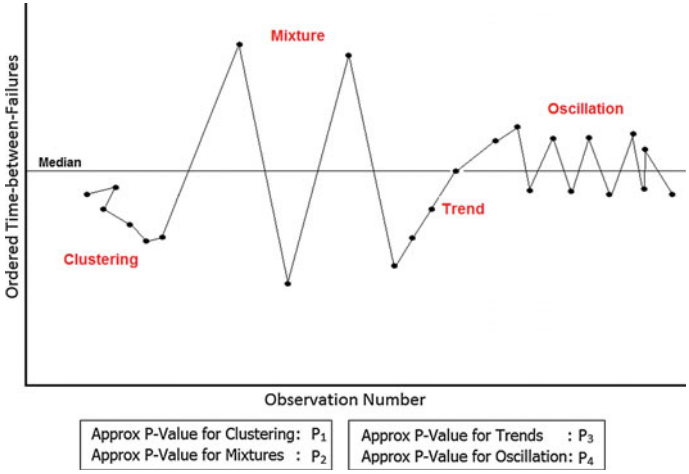


Fig. 2 Detection of special-cause data variations using run chart

These systems are called as non-stationary systems. Their reliability variations are required to be examined via stochastic processes. On the other hand, best-fit distributions are good enough for reliability assessment of non-trend systems. Data trend for individual components of dragline can be examined with hypothesis testing methods. In this respect, Crow-AMSAA and Laplace are effective hypothesis methods which tests the validity of homogenous Poisson process (HPP) in null hypothesis, in order to verify non-trend behavior. On the other hand, alternative hypothesis in the methods defense the validity of non-homogenous Poisson process (NHPP).

Crow-AMSAA test rejects the null-hypothesis in case $2N/\hat{\beta} < \chi^2_{2N,1-\alpha/2}$ or $2N/\hat{\beta} > \chi^2_{2N,\alpha/2}$, where χ and α are chi-squared distribution and confidence interval, respectively. In test statistic, N is total number of failure and $\hat{\beta}$ is called as shape parameter. $\hat{\beta}$ parameter can be estimated as in Eq. 1 [3]. In the equation, T_i and T_N are cumulative time-between-failures at ith failure and at Nth failure, respectively.

$$\hat{\beta} = \frac{N}{\sum_{i=1}^{N-1} \ln\left(\frac{T_N}{T_i}\right)} \tag{1}$$

On the other hand, Laplace test rejects validity of HPP when $U_L > z_{\alpha/2}$ and $U_L < -z_{\alpha/2}$. Test statistic, U_L, can be estimated using Eq. 2 [3]. N and T_i are same as in Crow-AMSAA test.

$$U_L = \frac{\sum_{i=1}^{N-1} T_i - (N-1)\frac{T_N}{2}}{T_N\sqrt{\frac{N-1}{12}}} \tag{2}$$

According to the tests, dragging chain assembly, hoisting brake, rigging socket, and rotation mechanism for Dragline-1 and hoisting rope-mode01, hoisting socket, bucket chain assembly, bucket pins, bucket ringbolt, rigging pulley-mode02, generator set, lubrication mechanism, rotation, and warning mechanisms for Dragline-2 were detected to exhibit trend behavior in their lifetime datasets. Reliabilities of these components were discussed using general renewal process which can evaluate the lifetime trend using a restoration factor. This process can be used with one of two approaches on virtual age of system. One approach assumes that maintenance recovers the deterioration only between the last and previous maintenance where the other approach assumes that cumulative deterioration can be recovered proportionally in maintenance activities. These models are called as Kijima-I and Kijima-II, respectively. This study utilizes Kijima-II assumption in the analyses, given in Eqs. 3–5 [4].

$$f(t_i|t_{i-1}, t_i, \ldots, t_1) = f(t_i|t_{i-1}) \tag{3}$$

$$\upsilon_i = q(\upsilon_{i-1} + x_i) \tag{4}$$

$$f(t_i|t_{i-1}) = \lambda\beta(x_i + \upsilon_{i-1})^{\beta-1} e^{-\lambda[(x_i + \upsilon_{i-1})^{\beta} - \upsilon_{i-1}^{\beta}} \tag{5}$$

In Eqs. 3–5, β is shape parameter, λ is failure rate, υ is virtual age, q is degree of repair, and x is time between failures. The model assumes that a system gets old differently from the calendar age, depending on the effectiveness of maintenance. System age can be called as virtual age. In case maintenance is carried out perfectly, virtual age remains same and stops to wear. In the process, effectiveness of maintenance is measured with degree of repair and it takes a value between 0 and 1 which are the limit values indicating perfect and minimal maintenance, respectively. This value can be utilized alternatively with restoration factor (q = 1−RF). Failure rate of a system with general renewal process can be estimated as in Eq. 6 [4]. Best estimate of the shape parameter, $\hat{\beta}$, can be calculated with Eq. 1 if the data set is failure truncated which means the observation is stopped at predetermined number of failure instead of predetermined time.

$$\hat{\lambda} = \frac{n}{T^{\hat{\beta}}} \tag{6}$$

Differently from trend-components, reliabilities of the other components with non-trend data behavior were assessed using best-fit distributions such as, Weibull, exponential, log-normal, and log-logistic. Considering the trend assumptions discussed above, time-dependent reliabilities of all components for both draglines were estimated using Reliasoft Weibull ++7 software. Since the components in each subsystem are connected to each other serially, failure rates of subsystems were obtained as in Figs. 3 and 4. The figures show that bucket and dragging units exhibit the most failure-intensive behavior where boom is the least-failure inducing unit for both draglines. These lifetime values will be utilized to describe random behavior of component failures in maintenance simulation in Sect. 3.

3 Simulation of the Maintenance Policy

Optimization of a maintenance policy requires a comparable model to measure the effectiveness of optimization over the current policy. Therefore, expected breakdown consequences of the current maintenance was modelled using random behavior of dragline components and deterministic behavior of inspections and compulsory halts in dragline operations. In the study, the analyses used objective data covering descriptive information of failures recorded during maintenance activities and subjective data achieved from dragline catalogues, maintenance crew,

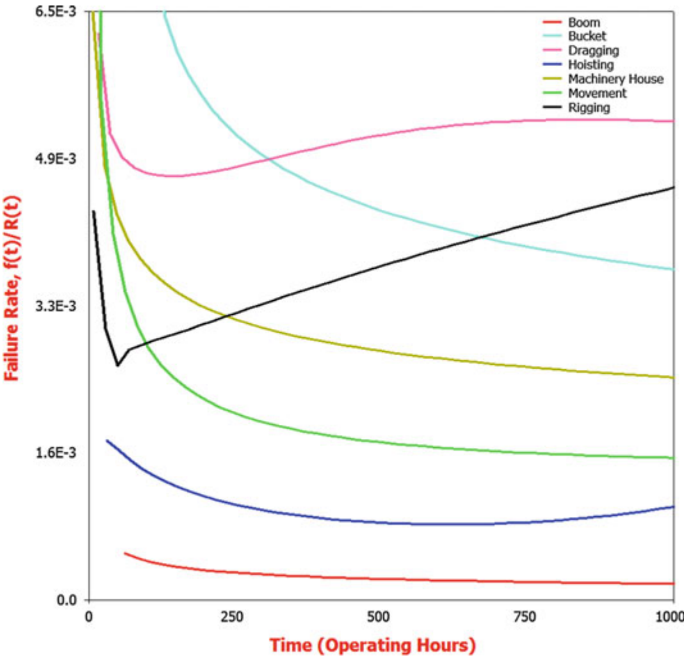


Fig. 3 Failure rates of Dragline-1 subsystems

and literature information. General maintenance and operation profiles of the draglines in Tunçbilek coal mine are given as follows:

i. Draglines perform overburden operations continuously all the year round. Daily work hours are divided into 3 shifts with 8 h intervals.
ii. Operation of the draglines is stopped compulsory for 30 min in each shift regarding employee rights. Therefore, utilization times of draglines are 22.5 h a day.
iii. Dragline operations are also stopped due to: (a) component failures, (b) regular inspections, (c) interruptions on energy transmission line, (d) unfavorable weather conditions, and (e) lack of sufficient maintenance staff.
iv. There are two common maintenance approaches for the draglines in the mine: (i) Corrective recovery of the components after failure and (ii) Performing 8-h regular inspections every 160 h.

During the simulations, the components were allowed to fail randomly according to their lifetime parameters. After each failure, components were assumed to be restored according to restoration factor estimated using general renewal process. For the components without any lifetime trend, maintenance activities were assumed to recover the component to as good as new condition. On the other hand, the components with lifetime trend were assumed to be restored to a condition between as good as new and as bad as old. In addition, regular inspections with 8 h

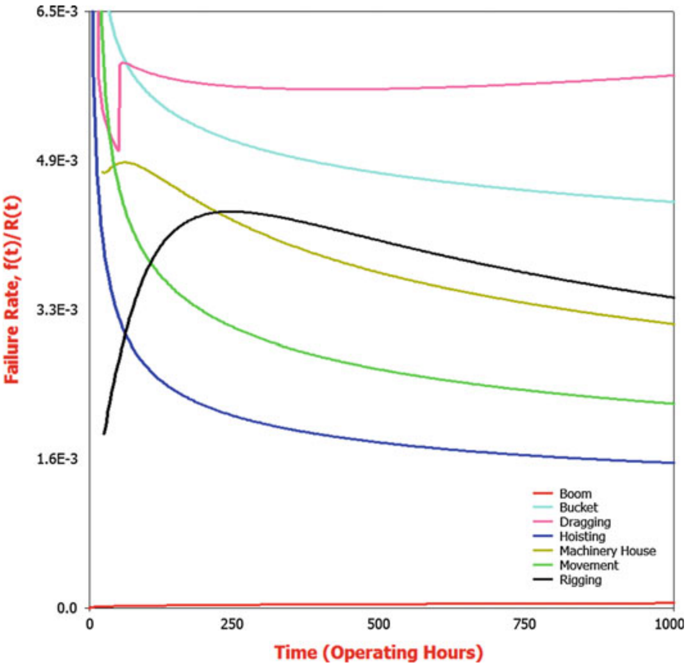


Fig. 4 Failure rates of Dragline-2 subsystems

and compulsory breaks in shifts with 30 min were also introduced in the simulation. These compulsory breaks cause halts of the draglines and pause virtual ages of the components during the events. Moreover, energy source statistics were also considered in the simulation to create more realistic policy model. The model was simulated with 1000 iterations using Reliasoft Blocksim 7 software. A representative illustration of one iteration covering major breakdowns can be viewed in Fig. 5.

The outputs of the simulations are stated in Table 2. The results reveal that the estimated availability of Dragline-1 and Dragline-2 is 64 and 69 % where 158 and 162 numbers of failures are expected to occur in a year, respectively. Moreover, occurrence of the failures and compulsory breaks are expected to cause 3164 and 2720 h of system halts for Dragline-1 and Dragline-2, respectively.

4 Future Study

Section 1–3 provides a basis in construction of maintenance optimization model for the draglines. In the future study, direct and indirect economic consequences of failures and breakdowns will be introduced to both the current and optimized model

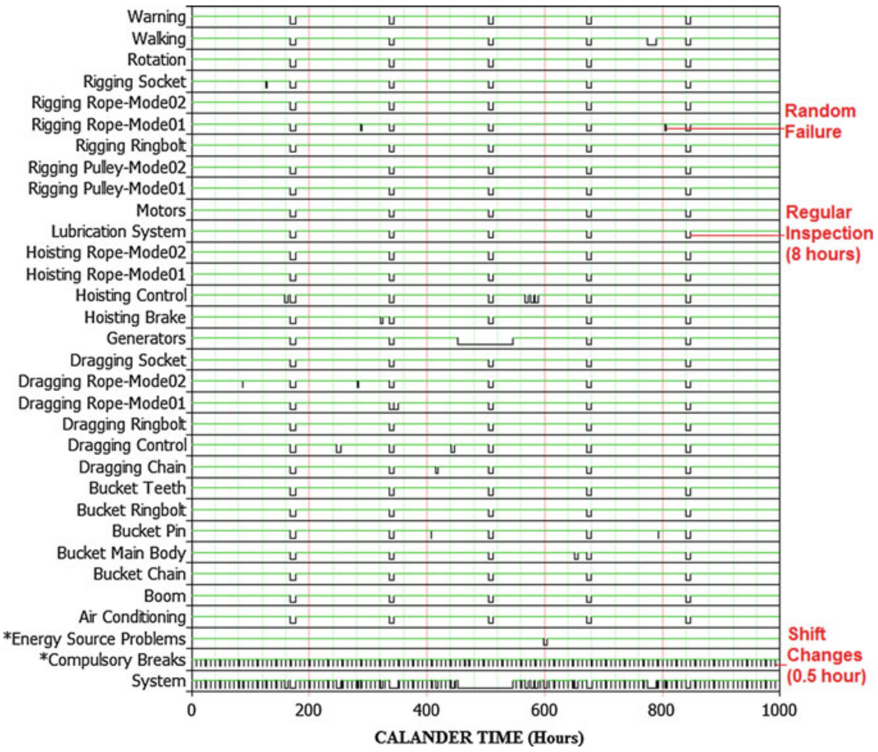


Fig. 5 Simulation window for one representative iteration

Table 2 Simulation results of the draglines

	Dragline-1	Dragline-2
General profile		
Mean availability	0.64	0.69
Std. deviation (mean availability)	0.04	0.03
Expected number of failures	158.05	161.54
Std. deviation (number of failures)	12.48	11.53
System uptime/downtime (hours)		
Uptime	5601.69	6045.92
Total downtime	3164.31	2720.08
Total duration	8766.00	8766.00
System downing events (number)		
Number of corrective maintenance	158.05	161.54
Number of inspections	52.00	52.80
Total events	210.05	214.34

to create more realistic and comparable model. Cost equations that will be utilized for estimation of these measures can be viewed in Eqs. 7–8.

$$\text{Unit Failure Cost} = \mathrm{C}_{\text{RepairDirect}} + \mathrm{C}_{\text{RepairIndirect}} \tag{7}$$

$$\mathrm{C}_{\text{Repair Indirect}} = MTTR_{component} \times \frac{V_{bucket} \times F}{S} \times \frac{1}{\frac{T_{cycle}}{\eta_{operator}}} \times C_{per\,bank\,m^3} \tag{8}$$

In Eq. 7, direct cost is physical consequence of a failure that can change component to component. It can cover the costs required for spare parts, hourly rate of crew, machine hiring, and energy consumption. On the other hand, indirect cost is non-physical cost of breakdown that can include production losses, penalty of unmet commitments, and damage in corporate image. Since production loss is the most measurable indirect cost, only this cost will be taken in calculations. Factors of production loss estimation can be viewed in Eq. 8. In the formula, *MTTR* is mean time to repair for components, V_{bucket} is volume of dragline buckets, *F* is fill factor, *S* is swell factor, T_{cycle} is cycle time, $\eta_{operator}$ is efficiency of operator in cycles, and $C_{per\,bank\,m^3}$ is the revenue for excavation of unit volume of bank material.

Including cost factors, the current policy will be improved via optimization tools such as age replacement decisions of wear-out components, opportunistic maintenance which aims simultaneous maintenance of same-duty components, and optimization of inspection intervals. Consequently, validity of the optimized policy will be discussed with a cost effectiveness analysis.

5 Conclusions

This study aims to simulate the maintenance policy of two draglines currently operating in Tunçbilek coal mine, in order to create a basis for the future maintenance optimization study. In this sense, lifetime characteristics of the components were achieved following the pre-processing of each lifetime dataset. These lifetime parameters were utilized to describe random failure profile of draglines. In addition, the compulsory halts in regular inspections and shifts were also included deterministically in the simulation. The results state that 158 and 162 numbers of failure are expected to take place annually for Dragline-1 and Dragline-2, respectively. The events in a year due to compulsory breaks and failure cause 3164 and 2720 h halting of Dragline-1 and Dragline-2 and yield availability of 64 and 69 %, respectively.

References

1. Ridley P, Algra R (2004) Dragline bucket and rigging dynamics. Mech Mach Theory 39:999–1016
2. Gilewicz P (1999) U.S. dragline census. Coal Age 104(8):35–40
3. Wang P, Coit DW (2005) Repairable systems reliability trend tests and evaluation. In: Proceedings of IEEE reliability and maintainability symposium
4. Mettas A, Zhao W (2005) Modeling and analysis of repairable systems with general repair. In: Proceedings of IEEE reliability and maintainability symposium

Reliability Analysis of Motor System of Dump Truck for Maintenance Management

Zeynab Allahkarami, Ahmad Reza Sayadi and Amol Lanke

Abstract Dump truck is one of the main machinery in open pit mines. From an economic point of view, more than 50–60 % of production costs in open pit mines are allocated to hauling and loading costs, so it is important to keep equipment in good condition. Reliability is a useful tool for evaluating the performance of this machine. In this research, the reliability of motor subsystem of a dump truck in Miduk Copper Mine in Iran has been analyzed. The failure data were collected during 20 months of dump truck operation. Trend and serial correlation tests were used to validate the assumption of independent and identically distribution (IID). According to tests, the data are independent and identically distributed therefore the renewal process technique is used for modelling. For finding the best-fit distribution, different types of statistical distributions were tested using the Easyfit software. The analysis results indicated the time between failures (TBF) data obey the Weibull (3p) distribution. The developed model based on these data showed that the reliability of the motor subsystem decreases to a zero value after approximately 430 h of operation. Regarding to the obtained reliability plot, preventive reliability-based maintenance time interval for 90 % reliability levels for machine in the motor subsystem is 21 h.

Keywords Reliability · Dump truck · Motor system · Renewal process

Z. Allahkarami · A.R. Sayadi (✉)
Faculty of Engineering, Department of Mining Engineering,
University of Tarbiat Modares, Tehran, Iran
e-mail: sayadi@modares.ac.ir

Z. Allahkarami
e-mail: z.allahkarami@modares.ac.ir

A. Lanke
Division of Operation and Maintenance Engineering,
Luleå University of Technology, Luleå, Sweden
e-mail: amol.lanke@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_50

1 Introduction

Reliability analysis is an important implement to assess the efficiency of a system and choose a maintenance strategy [1]. Reliability has been an accepted performance factor of systems. In the commercial area, high levels of reliability are also crucial [2]. In the open pit mining project, equipment such as dump trucks play key role in the production plan, so their performance is very imperative for engineers and managers. In modern mining the forecasting production amount is essential for managers and stakeholders. A common reason why the production amount is not according to plan, is the unavailability of equipment, and one of the main type of machines in most open pit mines is the trucks [3]. Kumar [4] modelled the reliability of load-haul-dump (LHD) fleet. Hall and Daneshmend [5] analyzed reliability and maintainability of mobile underground haulage equipment for improving decision making for maintenance planning. Barabady and Kumar [6] studied the reliability of a crushing plant and identified the critical subsystems, then showed that reliability study is very valuable for maintenance scheduling. Hoseinie et al. [7] analyzed reliability of the shearer machine to detect critical subsystems. Then in order to achieve a proper and practicable maintenance schedule, a task package was suggested for the drum shearer machine in the Tabas coal mine. Morad et al. [8] assessed the reliability of 10 trucks and computed importance of each component was by weighted importance measure method. This study showed the impact of critical items on the availability of machines. Dump truck is one of the main machinery in open pit mines and its downtime reduction has direct effects on production plan. In this paper, a dump truck of Miduk copper mine in Iran with the age of approximately 15,600 operation hours was considered. Technical specifications of the dump truck are listed in Table 1. The maximum and minimum

Table 1 Specifications of HD875-5

Specifications of HD875-5	
Engine	
Gross horsepower	783 kW/1050 HP
Flywheel horsepower (SAE J1349)	753 kW/1010 HP
Capacities	
Heaped (2:1 SAE)	40 m^3
Payload maximum	60 m^3
Maximum gross vehicle weight	166,000 kg
Body	
Floor	19 mm
Front	12 mm
Sides	9 mm
Other	
Max. travel speed	65 km/h
Min. turning radius	9.9 m

temperature in this region is +35 and −15 °C respectively. After analysis of number of failure, it was determined motor sub-system of the dump truck has the top failure frequency, so this sub-system was selected for reliability analysis.

2 Reliability

Reliability is viewed as both an engineering and a probabilistic concept [9] and it is defined as “the duration or probability of failure-free performance under stated conditions” [2]. The reliability function, R(t), or the probability of a system not failing prior to time t, is determined by [2, 3]:

$$R(t) = 1 - F(t) = 1 - \int_0^t f(t)dt \tag{1}$$

where R(t) is the reliability at time t; F(t) the cumulative failure distribution function; and f(t) the failure probability density function.

3 Data Gathering and Analysis

Data is the primary foundation for statistical reliability analysis. Accordingly, failure data of a dump truck in a 20-month period in Miduk copper mine was gathered and dump truck was represented into 6 subsystems (Fig. 1).

Data showed the most frequent failure event was in the motor system. Motor has a major effect on this truck downtime. Therefore, a concentration on reliability of motor is important to improve the performance of the machine. The result is shown in Fig. 2.

The motor subsystem of dump truck consists of following parts [8];

- Engine body
- Cooling
- Lubrication
- Intake and exhaust
- Fuel

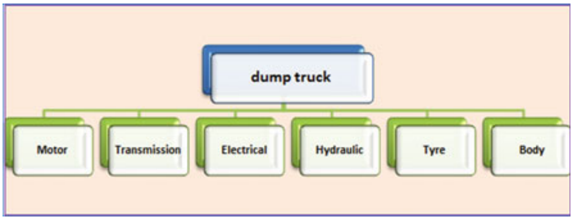


Fig. 1 Dump truck subsystems

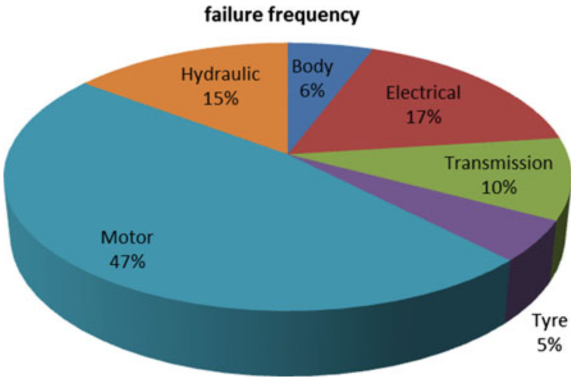


Fig. 2 Failure frequency of dump truck subsystems

The failures of these components were considered as motor subsystem failures. Main reasons of motor downtime are intake and exhaust failures. Other subsystems like the hydraulic is a composite of hydraulic tank, brake control, hoses, pipes, joints, hoist cylinders and related items. The electrical components are battery, alternator, cable, starter and lights. Gearbox, differential, universal joint, clutch and wheels were considered as Transmission subsystem.

For choosing an appropriate methodology of modelling, two statistical tests must be done to validate the assumption of independent and identically distributed (IID) for the time between failure (TBF) data. The analysis procedure is shown in Fig. 3.

There are graphical and analytical approach for testing presence of trends and serial correlation. In this paper, graphical methods have been used, because of simplicity, quick performance and their valuable information. A scatter plot of cumulative TBFs versus cumulative number of failures is used to test the presence of trend, if the plot is approximately a straight line, then the TBF data is identically distributed and free from trends [10]. Figure 4 shows the trend test on the motor subsystem of dump truck. Military Handbook Test is one of the analytic trend tests, that is applied in this paper, too. This test using calculating the test statistic U (Eq. 2) checks the trend of data [11]:

$$U = 2\sum\nolimits_{i=1}^{n-1} \ln(T_n/T_i) \tag{2}$$

where:
n: is total number of failure,
Tn: is time of the nth failure,
Ti: is time of the ith failure.

The test statistic U is chi-squared distributed with 2(n−1) degrees of freedom under the null hypothesis of an HPP [9, 10]. Result of analytic trend test is presented in Table 2.

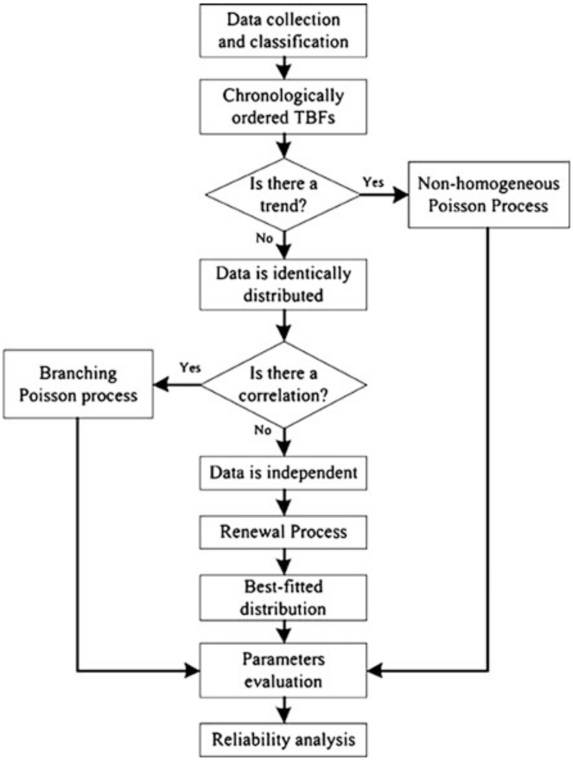

Data collection and classification
Chronologically ordered TBFs
Is there a trend?
Yes
Non-homogeneous Poisson Process
No
Data is identically distributed
Is there a correlation?
Yes
Branching Poisson process
No
Data is independent
Renewal Process
Best-fitted distribution
Parameters evaluation
Reliability analysis

Fig. 3 The reliability analysis procedure [6, 7]

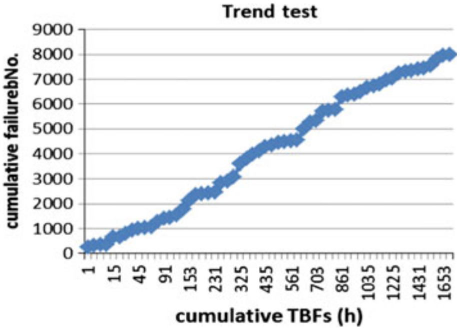


Fig. 4 Trend test on motor system of dump truck

Serial correlation can be tested by plotting the (i)th TBF against (i-1)th TBF. If the plot is without any patterns, TBF data are serial correlation-free [7, 12, 13] (Fig. 5).

The results of tests on TBF data show the data are independent and identically distributed (iid), so renewal process (RP) is the best method for modelling.

Table 2 The result of analytic trend test for TBF

Subsystem: motor				
Calculated statistic U	Critical value of chi-square distribution at 5 % level of significance	Degree of freedom	IID	Modelling method
114.23	90.35	114	Yes	RP

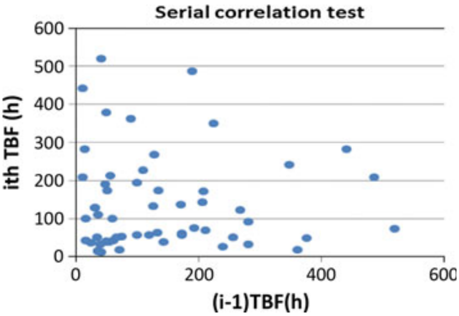


Fig. 5 Result of correlation test

4 Reliability Modelling

Easyfit software was used to find the proper distribution. Common distributions such as Weibull, Exponential, Lognormal and etc. distributions in reliability analysis were tested. Then Kolmogorov–Smirnov (K–S) test was applied in selecting the best distributions for reliability model. According to analysis, Weipull (3P) was the best of all. The result is illustrated in the Table 3.

The achieved reliability plot is presented in Fig. 6. The developed model based on these data shows that the reliability of the motor subsystem decreases to a zero value after approximately 430 h of operation and reliability decreases to 50 % after 90 h.

5 Maintenance

Reliability Cantered Maintenance (RCM) is an engineering method for determining the level of an organization's maintenance program and reliability is a key concept of it. In this paper, different level of reliability of the motor subsystem in the dump truck was considered for programing preventive maintenance (PM) intervals

Table 3 Parameters estimation for modelling

Subsystem	Mean time between failure (h)	Parameters
Motor	139.66	α: 883 β: 120.05 γ: 12

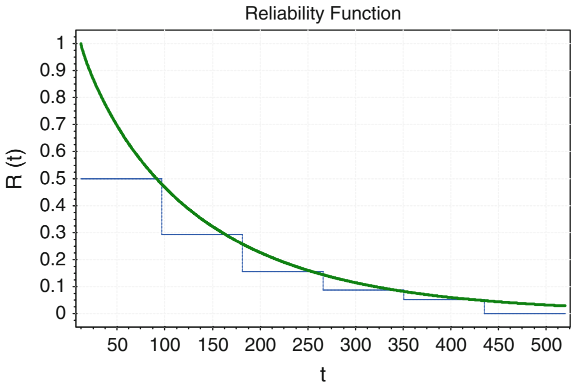


Fig. 6 Reliability plot of motor subsystem

Table 4 Preventive maintenance intervals for motor sub-system of the truck at different reliability levels

Reliability levels (%)	PM intervals (h)
90	21
80	34
70	49
60	68
50	91
5	428

(Table 4). One of the most important objectives of PM is improving the reliability of equipment. Using the predicted intervals in desirable level for maintenance managers can cause reduction of unpredicted downtime and costs. These time intervals are scheduled to inspect, correct, clean, lubricate, replace spare parts, and repair the system.

6 Conclusion

The reliability of main equipment in open pit such as dump trucks is essential and it has significant impact on production goals and maintenance costs. In this paper, data of a truck in Miduk mine in Iran were gathered and classified. Analysis of data showed the most frequent failure event was in the motor system of this dump truck, so it was focused on the motor subsystem for improving reliability of the dump truck. The most critical subsystems are those that have the most failure frequency. RCM recognizes the most critical components and optimizes their maintenance strategies using appropriate and cost-effective methods. Results of analysis showed that the TBFs data are IID, Hence, the renewal process was used for reliability analysis. The data obeyed Weibull (3p) distribution. Using reliability plot, maintenance intervals were predicted. These time intervals should be scheduled to inspect, correct, clean, lubricate, replace spare parts, and repair the system. It is obvious that managers and engineers can focus on PM intervals to improve reliability of the system.. For PM, the tasks can be grouped for executing in the most economical way. The improved reliability by RCM leads to fewer failures, more availability and lower maintenance costs.

References

1. Wang Z, Huang H-Z, Du L (2011) Reliability analysis on competitive failure processes under fuzzy degradation data. Appl Soft Comput 11(3):2964–2973
2. Handbook M (1988) Electronic reliability design handbook. MIL-HDBK-338, DoD
3. Dhillon BS (2008) Mining equipment reliability, maintainability, and safety. Springer, London
4. Kumar U (1990) Reliability analysis of load-haul-dump machines. Lulea University of Technology, Lulea
5. Hall RA, Daneshmend LK (1072) Reliability and maintainability models for mobile underground haulage equipment. CIM Bull 2003(96):159–165
6. Barabady J, Kumar U (2008) Reliability analysis of mining equipment: a case study of a crushing plant at Jajarm Bauxite mine in Iran. Reliab Eng Sys Saf 93(4):647–653
7. Hoseinie SH et al (2012) Reliability analysis of drum shearer machine at mechanized longwall mines. J Qual Maint Eng 18(1):98–119
8. Morad A, Pourgol-Mohammad M, Sattarvand J (2014) Application of reliability-centered maintenance for productivity improvement of open pit mining equipment: case study of Sungun Copper mine. J Cent S Univ 21(6):2372–2382
9. Modarres M, Kaminskiy MP, Krivstov V (1999) Reliability engineering and risk analysis: a practical guide. CRC press, Boca Raton
10. Kumar U, Klefsjö B, Granholm S (1989) Reliability investigation for a fleet of load haul dump machines in a Swedish mine. Reliab Eng Sys Saf 26(4):341–361
11. Standard M (1981) MIL-HDBK-189. Reliability growth management
12. Barabady J (2005) Reliability and maintainability analysis of crushing plants in Jajarm Bauxite mine of Iran. In: Proceedings—annual reliability and maintainability symposium. IEEE
13. Kumar U, Klefsjö B (1992) Reliability analysis of hydraulic systems of LHD machines using the power law process model. Reliab Eng Sys Saf 35(3):217–224

Part VIII
Software Reliability & Data Quality

Multi Up-Gradation Reliability Model for Open Source Software

Mahdieh Ahmadi, Iraj Mahdavi and A.H.S. Garmabaki

Abstract Nowadays, software companies have to continuously do up-gradation or add-ons in their software to survive in the market. This paper presents an effective reliability model for multi release open source software (OSS), which derived based on software lifecycle development process (SDLC) proposed by Jørgensen [1]. Most of OSS reliability models proposed in the literature are based on closed-form methodology and do not consider the properties of OSS in the model structure. The proposed model, incorporate bugs removed from two different phases, namely a pre-commit test and parallel debugging test. Furthermore, the proposed model is based on the assumptions that the overall fault removal of the new release depends on the reported faults from the previous release of the software and on the faults generated due to adding some new functionalities to the existing software system. The parameters of model have been estimated on real software failure dataset with three releases and goodness of fit of values have been calculated. Results show that the proposed model fits the data reasonably well and present better accuracy in comparison with other methods.

Keywords NHPP · Multi release · Up-gradation · Open source software (OSS) · Testing phase

M. Ahmadi · I. Mahdavi
Department of Industrial Engineering, Mazandaran University of Science and Technology, Babol, Iran
e-mail: mahdiehahmadi@yahoo.com

I. Mahdavi
e-mail: irajarash@rediff.com

A.H.S. Garmabaki (✉)
Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden
e-mail: amir.garmabaki@ltu.se; garmabaki@gmail.com

A.H.S. Garmabaki
Department of Mathematics and Computer Science, Islamic Azad University, Nour Branch, Nour, Iran

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_51

Notations

The following notation used in the paper

$m(t)$	The expected number of faults removed by time *t*.
$\lambda(t)$	Failure intensity.
$F(t)$,	Probability distribution functions for FRP.
$F_i^{PCT}(t)$, $F_i^{PDT}(t)$	Probability distribution functions for pre-commit test (PCT) and parallel debugging test (PDT), respectively.
$F_i^{PR}(t)$	Probability distribution functions for bugs reported from Production Release (PR) previous version.
τ_i	Time for *i*th release, $i = 1..n$.
a_i	Initial fault content for *i*th release, $i = 1..n$.
a	Total fault content in the software.
β_1, β_2, β_3	The shape parameter of the Weibull model for *i*th release during PCT, PDT and PR; $i = 1..n$.
θ_1, θ_2, θ_3	The scale parameter of the Weibull model for *i*th release during PCT, PDT and PR; $i = 1..n$.
λ	Proportion of fault removed by testing team during PCT.
$1-\lambda$	Proportion of fault removed based during PDT.

1 Introduction

Open source software is defined as the software whose source code is available along with the software and user has the freedom to distribute, run, copy, change, and improve the software under the licensing policies of OSS [2]. OSS methodology provides greater value to users and leads to increased revenue for the OSS companies. Many different developers, user, or co-developers can participate in the development of the OSS. The development of OSS is always initiated by a single developer or a single group, who starts the development of software for its own "personal itch" [3].

Software reliability is one of main performance measures for the quality of the software. Software reliability is defined as the probability of failure-free software operation for a specified period of time in a specified environment. A software reliability model (SRM) provides a mathematical relationship between time span of testing and the cumulative number of faults detected. Software testing involves running the software and checking for unexpected behavior in the software output. The process of locating the faults and designing the procedures to detect them is called the debugging process [4–7].

Software Reliability Growth based Models (SRGMs) are one of the successful model to describe software failure-occurrence or fault-detection phenomenon in the testing and operational phases. Several studies had been done for reliability analyzing of OSS [8–10]. For example, Eclipse, Apache HTTP Server 2, Firefox, MPlayer OS X, and ClamWin Free Antivirus applications have been evaluated by means of several models, and the Weibull distribution has been found to adapt well in modelling simpler projects, although more complex models are claimed to be needed for Firefox and Eclipse [8, 11]. In [12], several OSS projects have been analyzed and found that, the Weibull distribution has been found to be a simple and effective way to represent software reliability growth. In addition, they conclude that open source projects exhibit similar reliability growth pattern with that of closed source project. Rossi et al. [13] show that the Weibull model can be used for the reliability analysis of OSS successfully. For predictive ability, the Weibull model is definitely good to estimate the total number of failures but it cannot be used as any other SRGM for early prediction.

Due to time and resource limitation during testing phase, software companies do not attempt to deliver a complete and perfect software product in one development cycle. They plan successive releases of software by adding new features or new functionalities or try to improve the performance of the system as compared to previous releases. Mozilla Fire Fox, GNOME, Microsoft Windows and Office, Adobe, represent good examples of such practice. This strategy provides several benefits for software companies which discussed by Garmabaki et al. [4].

Upgrading a software application is a complex task. The upgraded and existing system may differ in the performance, interface and functionality etc. Although safe up-gradation can improve the behavior of the system and can preserve market for company, risky up-gradation can cause critical error in system.

In the useful-life cycle phase, software companies introduce new add-ons or features based on the user need. Hence, software will experience an increase in failure rate, each time an upgrade is made. The failure rate decreases gradually, because of the faults/failures found and fixed after the upgrades. Figure 1 depicts the increase in failure rate due to the addition of new features in the software.

Recently Singh et al. [10], Kapur et al. [6], Garmabaki et al. [4] developed multi up-gradation reliability model. The proposed model is based on the assumption that the overall fault removal of the new release depends on the reported faults from the

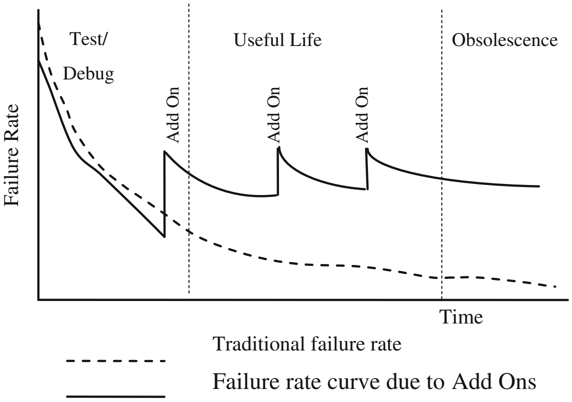


Fig. 1 Failure rate in classical SRGM and multiple releases [11]

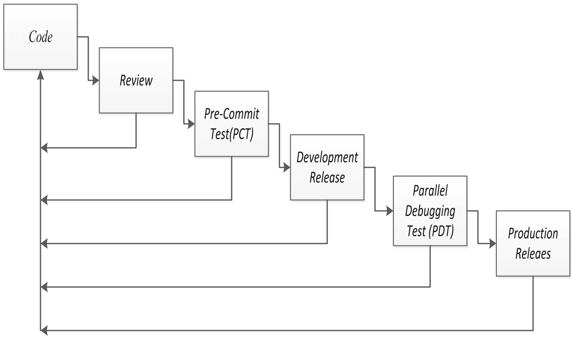


Fig. 2 Life cycle model for OSS [1]

just previous release of the software and on the faults generated due to adding some new functionalities (add-ons/up-gradations) to the existing software system.

In this paper, we incorporate bugs removed from pre-commit test and bugs reported by parallel debugging test based on SDLC proposed by Jørgensen [1] for OSS, which has been depicted in Fig. 2. Furthermore, the relation between fault removal processes (FRP) of successive release of the software are considered and shown in the Fig. 3. The rest of this paper is organized as follows. Section 2 discuses about Weibull model. In Sect. 3 briefly reviews Jørgensen [1] SDLC model. The proposed model is introduced in Sect. 4 and the relation between different testing phase and releases was discussed. Section 5 shows the experimental results through real data sets and estimation of parameters. Finally, summery and conclusions are given in Sect. 6.

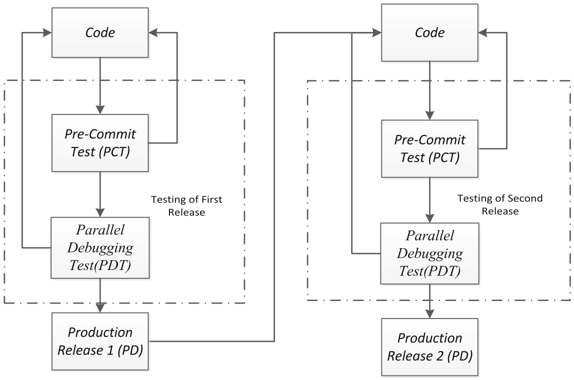

Code
Pre-Commit Test (PCT)
Testing of First Release
Parallel Debugging Test(PDT)
Production Release 1 (PD)
Code
Pre-Commit Test (PCT)
Testing of Second Release
Parallel Debugging Test(PDT)
Production Release 2 (PD)

Fig. 3 Testing process for multi release OSS

2 Weibull Model

For the past three decades, various mathematical models have been proposed to assess the software reliability. The non-homogeneous Poisson process (NHPP) based SRGMs have proved quite successful in practical software reliability engineering [7]. The main issue with the NHPP model is to determine an appropriate mean value function to denote the expected number of failures experienced up to a certain time point.

Let $\{N(t),\ t \geq 0\}$ be a counting process representing the cumulative number of software faults removed by time *t*. The counting process $\{N(t),\ t \geq 0\}$ is shown to be an NHPP with a mean value function *m*(*t*), which represents the expected number of faults removed up to a certain time.

Based on the NHPP assumption, it can be shown that $\{N(t),\ t \geq 0\}$ has a Poisson distribution with mean *m*(*t*), i.e.,

$$Pr\{N(t) = n\} = \frac{m(t)^n}{n!}.\mathrm{e}^{(-m(t))}, \quad n = 0, 1, 2, \ldots \tag{1}$$

By definition, the mean value function of cumulative number of faults, *m*(*t*), can be expressed in terms of the failure intensity function of the software, i.e,

$$m(t) = \int_0^t \lambda(s)ds = a.F(t) \tag{2}$$

where *F*(*t*) is cumulative distribution function for fault removal times [5, 14].

Unlike traditional closed source software, OSS involves much more testers in the testing process and most of these testers are contributors. The number of contributors involved in the OSS is largely influenced by the attractiveness of the software [3]. More specifically, the adaptation behavior of users/contributors is increasing over a certain point of time and after that decrease since the software is losing its attractiveness. This characteristic for OSS reflects an initial increase and eventual decrease in fault/failure occurrences [15] and the Weibull model is flexible enough to capture this behavior. In addition, the Weibull model empirically can fit many types of failure data, especially for OSS.

The mean number of faults removed by Weibull model is given as:

$$m(t) = a \cdot \left(1 - e^{-\left(\frac{t}{\theta}\right)^{\beta}}\right) = a \cdot F(t); \quad \beta > 0,\, \theta > 0 \tag{3}$$

where

$$F(t) = \left(1 - e^{-\left(\frac{t}{\theta}\right)^{\beta}}\right) \tag{4}$$

Is cumulative Weibull distribution.

3 Testing Phase in Jørgensen [1] SDLC

The software development life cycle (SDLC) for OSS is different from the traditional commercial software called as "closed software". It is found by various researchers that the traditional SDLC cannot be used for the development of OSS [16]. Various researchers and practitioners are working on developing the standard development life cycle of OSS.

Software testing is very important phases in the SDLCs and it has direct influence on operational phase and reliability of software. Software testing is defined as the process of executing a program to locate an error [17].

Jørgensen [1] provides a life cycle model, shown in Fig. 2. This model is widely accepted as a framework for the OSS development. In this model, Software testing is carried out in the following two phases, which is defined as:

Pre-Commit test: The reviewed code, then passed through an unstructured testing phase. A pre-commit check is invoked right before a change is committed into the repository. The developed code is tested to find errors and un-necessary code is rejected during this phase. The commit operation is performed on the code that is found necessary and accurate. This phase is considered most important in the development process because if not performed properly, and it may lead to failure for the OSS.

Parallel Debugging: Once the development has been released the code is exposed to a large number of contributors or user. They perform rigorous debugging to find all bugs and report them to the core developers.

4 Multiple *Release* Model

The proposed model in this paper is based on the assumption that the overall fault removal consists of

- Bugs removed from pre-commit test and bugs reported by parallel debugging test due to adding some new functionality (add-ons/up-gradations) to the existing software system.
- Bug reported from previous release to the testing team of the new release of the software.

The relation between different phases in teasing phase for OSS with successive releases is depicted in Fig. 3.

The basic assumptions of the model being proposed in the paper are as follows:

Assumptions:

(1) The FRP for each release is modeled by non-homogeneous poison process (NHPP).

(2) The number of faults detected at any time is proportional to the remaining number of faults in the software.
(3) The FRP incorporates bugs removed from pre-commit test and bugs reported by parallel debugging test.
(4) The undetected faults of previous release are removed by fault reported bug in the current release of the software.
(5) The number of faults in the beginning of the testing phase is finite.
(6) All faults are mutually independent from failure detection point of view.
(7) Each time a failure occurs, the error that caused it is immediately fixed, and no other errors are introduced.

In practice, it is important to know that how many faults exist in the software at any time, so that different testing strategy and testing effort can be applied to remove those faults.

Let consider that testing begins at time $t = 0$ and the first release of software be done at $t = \tau_1$. Note that we can't remove all faults during the testing phase and some of the fault remain in the code even after its release. The mathematical equation of these finite fault count model is given as:

$$\begin{aligned} m_1(t) &= a_1.\left(\lambda_1 F_1^{DCT}(t) + (1-\lambda_1)F_1^{PDT}(t)\right) \\ &= a.\, G_1(t);\ 0 < t \le \tau_1 \end{aligned} \tag{5}$$

where $G_1(t) = \lambda_1 F_1^{DCT}(t) + (1-\lambda_1)F_1^{PDT}(t)$.

Note that during the testing phase of first release no bug report is available. Thus the faults are removed on the basis of testing only.

After each release, one of the major issues faced by the software companies is to determine what new functionality/feature should be added in the next release of their product for surviving in the market because of competition. At this time company has information about bugs reported from the users of first release, which are in the operational phase. On the basis of these bug reports and market feedback, the company adds some new functionality to the existing software system or increase performance by removing bugs from the current system. Adding some new functionality to the software lead to increase in the fault content. At this stage, the model distinguishes between removal process related to the faults of the new code and undetected fault of previous release. During testing, it is quite possible that some faults of old code are removed directly by the testing team of the new release (without using any bug reports of the previous release of the software) and some others are removed on the basis of reported bugs. In addition, due to parallel testing, several people may report the same fault/failure which we call duplicates. Duplicates are not included in count of unique fault. Based on the above framework, we can write following mathematical equation for a second release.

$$m_2(t) = a_2.G_2(t - \tau_1) + (a_1 - m_1(\tau_1)).F_2^{PR}(t - \tau_1) \tag{6}$$

where $G_2(t - \tau_1) = \lambda_2 F_2^{DCT}(t - \tau_1) + (1 - \lambda_2)F_2^{PDT}(t - \tau_1)\ \tau_1 < t \leq \tau_2$.

At this step, faults generated due to the enhancement of the features are removed with $a_2.G_2(t - \tau_1)$ and $(a_1 - m_1(\tau_1))$. represent undetected faults of the first release, which interacts with the new portion of code and are removed/detected by testing team of the second release. i.e. $\mathrm{F}_2^{PR}(t - \tau_1)$.

The same situation will happen on the ith version at time t = τ_i as given by:

$$m_i(t) = a_i.G_i(t - \tau_{i-1}) + (a_{i-1} - m_{i-1}(\tau_{i-1} - \tau_{i-2})).F_i^{PR}(t - \tau_{i-1}) \tag{7}$$

$$\tau_{i-1} < t \leq \tau_i$$

where $G_i(t - \tau_{i-1}) = \lambda_i F_i^{DCT}(t - \tau_{i-1}) + (1 - \lambda_i)F_i^{PDT}(t - \tau_{i-1})$, $F_i^{DGT}(t - \tau_{i-1})$, $F_i^{PDT}(t - \tau_{i-1})$ and $F_i^{PR}(t - \tau_{i-1})$ are Weibull distribution for DCT, PDT, and PR, respectively.

5 Numerical Examples and Model Evaluation

Real software data sets from a famous open source project, namely as Apache 2.0.35 (first release), Apache 2.0.36 (second release), and Apache 2.0.39 (third release), are selected to validate the proposed model [18]. The Apache Software Foundation is a well-organized community, which developers cooperating under OSS methodology. The community's large size makes it a state-of-the-art community in terms of management of OSS projects.

To compare the proposed model with traditional reliability assessment models, we select the widely used Goel and Okumoto [12], Yamada et al. [19] and Li et al. [18] models.

We use Least Squares Method, one of the most significant methods to estimate model parameters in the software reliability field. For parameter estimation, we apply a statistical package for social science, "SPSS" software. We estimate the parameters a, θ_1, β_1, θ_2, β_2 and λ_i from the data sets. The estimated values of the parameters for the model in each data set are given in Table 1.

Table 2 shows the comparison criterion values related to the proposed model, along with recent and traditional software reliability models. The two important criteria used in the paper are defined as:

- Coefficient of Multiple Determination (R^2):
  We define this coefficient as the ratio of the sum of squares resulting from the trend model to that from the constant model subtracted from 1, i.e.

Table 1 Parameter estimates of proposed model

	Releases i		
Parameter	1	2	3
a	74	50	61
θ_1	6.05	12.6	29.4
β_1	0.89	0.99	1.4
θ_2	18.09	34.3	12.4
β_2	2.07	9.1	0.5
β_3	–	1.23	0.94
θ_3	–	7.1	9.2
λ_i	0.37	0.62	0.83

Table 2 Comparison criteria of proposed model

		MSE	Ad-R^2
First release	Proposed model	2.16	0.996
	Go-model	5.76	0.987
	Yamada model	7.7	0.983
	Xiang Li model	2.8	0.992
Second release	Proposed model	0.51	0.997
	Go-model	8.84	0.953
	Yamada model	8.39	0.955
	Xiang Li model	5.98	0.985
Third release	Proposed model	0.75	0.995
	Go-model	2.57	0.989
	Yamada model	2.4	0.988
	Xiang Li model	0.68	0.995

$$\mathrm{R}^2 = 1 - \frac{\text{residual SS}}{\text{corrected SS}} \tag{8}$$

- Mean Square Error (MSE):
  MSE is the difference between the expected values, $\hat{m}(t_i)$ and the observed data, y_i, measured as follows:

$$MSE = \sum_{i=1}^{k} \frac{(\hat{m}(t_i) - y_i)^2}{k} \tag{9}$$

  where k is the number of observations. A lower MSE indicates the less fitting error, thus better goodness of fit.

Figures 4, 5 and 6 show the estimated and actual values of the number of faults removed for each release, separately.

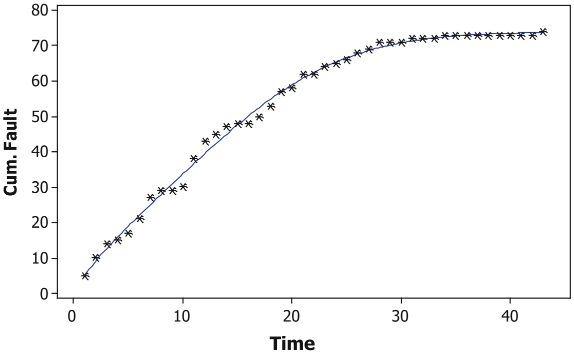


Fig. 4 Goodness of fit of first release

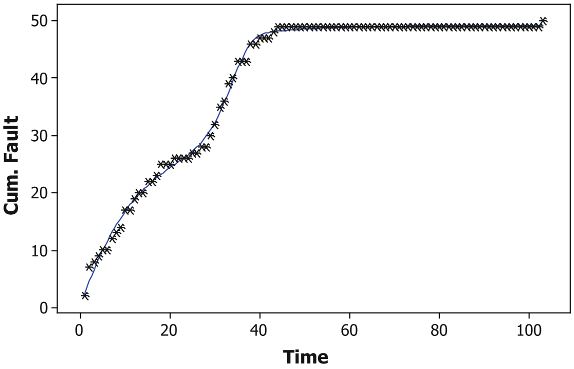


Fig. 5 Goodness of fit of second release

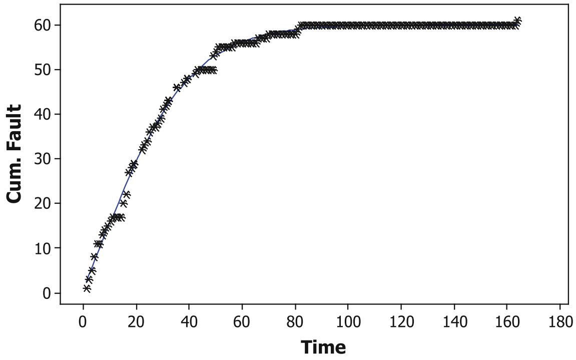


Fig. 6 Goodness of fit of the third release

6 Summery and Conclusion

In this paper, reliability modeling of successive release of OSS is investigated. During the testing process of OSS development, unlike closed-form software, large numbers of testers are involved and most are volunteers/contributors (free duty).

Utility and attractiveness of the software are two main issues that influence the number of contributors in OSS projects. More specifically, OSS can attract an increasing number of contributors in early stages, but after the number of contributors reaches a peak, it will decrease as the software loses its attractiveness over time. Hence, the mixture-type Weibull model is used to capture the cumulative fault removed of OSS. The proposed model is based on the assumptions that the overall fault removal of the new release depends on the reported faults from the previous release of the software and on the faults generated due to adding some new functionalities to the existing software system. The model validation is done and the parameters of model have been estimated on real software failure dataset with three releases and goodness of fit of values has been calculated. Furthermore, Goel and Okumoto [12], Yamada et al. [19] and Li et al. [18] models are selected to compare with the proposed model. Result show that the proposed model fits the data reasonably well and present better accuracy in comparing with other methods.

References

1. Jørgensen N (2001) Putting it all in the trunk: incremental software development in the FreeBSD open source project. Inf Sys J 11:321–336
2. Amant KS, Still B (2009) Handbook of research on open source software—technological, economic, and social perspectives
3. Raymond ES (2001) The cathedral & the bazaar: musings on Linux and open source by an accidental revolutionary. O'Reilly Media, Inc., California
4. Garmabaki AHS, Aggarwal AG, Kapur PK, Yadavali VSS (2012) Modeling two-dimensional software multi-upgradation and related release problem (a multi-attribute utility approach). Int J Reliab Qual Saf Eng 19:1250012
5. Kapur PK, Pham H, Gupta A, Jha PC (2011) Software reliability assessment with OR applications. Springer, London
6. Kapur PK, Singh O, Garmabaki AS, Singh J (2010) Multi up-gradation software reliability growth model with imperfect debugging, Int J Sys Assur Eng Manag 1:299–306
7. Musa JD, Iannino A, Okumoto K (1987) Software reliability: measurement, prediction, application. McGraw-Hill, NewYork
8. Rahmani C, Siy H, Azadmanesh A (2009) An experimental analysis of open source software reliability. Department of Defense/Air Force Office of Scientific Research
9. Tamura Y, Yamada S (2009) Optimisation analysis for reliability assessment based on stochastic differential equation modelling for open source software. Int J Syst Sci 40:429–438
10. Singh V, Kapur P, Tandon A (2010) Measuring reliability growth of open source software by applying stochastic differential equations in second world congress on software engineering (WCSE), pp 115–118
11. Kapur P, Pham H, Gupta A, Jha P (2011) Software reliability assessment with OR applications: Springer, London
12. Goel AL, Okumoto K (1979) Time-dependent error-detection rate model for software reliability and other performance measures. IEEE Trans Reliab 28:206–211
13. Rossi B, Russo B, Succi G (2010) Modelling failures occurrences of open source software with reliability growth, in open source software: new horizons, Springer, London, pp 268–280
14. Ohba M (1984) Software reliability analysis models. IBM J Res Dev 28:428–443
15. Garmabaki AH, Kapur P, Aggarwal AG, Yadavali V (2014) The impact of bugs reported from operational phase on successive software releases. Int J Prod Qual Manag 14:423–440

16. Pressman RS (2005) Software engineering: a practitioner's approach. Palgrave Macmillan, London
17. Pham H (2006) System software reliability. Springer, Berlin
18. Li X, Li YF, Xie M, Ng SH (2011) Reliability analysis and optimal version-updating for open source software. Inf Softw Technol 53:929–936
19. Yamada S, Ohba M, Osaki S (1983) S-shaped reliability growth modeling for software error detection. IEEE Trans Reliab 32:475–484

Quality of Streaming Data in Condition Monitoring Using ISO 8000

Mustafa Aljumaili, Ramin Karim and Phillip Tretten

Abstract The purpose of this paper is to propose a Data Quality Measurement Model based on ISO 8000 standard. This paper deals about the concepts implied in the measurement process, not about the measures themselves. Poor quality information causes customer dissatisfaction, lost revenue and higher costs associated with additional time to reconcile information. An understanding of the characteristics of the data that determine its quality, and an ability to measure, manage and report on data quality is required. Measurement is a major activity in data quality management. In literature, there are many proposals contributing somehow to the measurement of data quality. However, these measurement methods lack the unification. ISO 8000 provides a framework for improving data quality that can be used independently or in conjunction with quality management systems. ISO 8000 defines characteristics that can be tested by any organization in the data supply chain to objectively determine conformance of the data to ISO 8000.

Keywords Data quality · Information · eMaintenance · ISO 8000 · XML · Measurement · Assessment

M. Aljumaili (✉) · R. Karim · P. Tretten
Division of Operation, Maintenance and Acoustics, Luleå University of Technology,
971 87 Luleå, Sweden
e-mail: mustafa.aljumaili@ltu.se

R. Karim
e-mail: ramin.karim@ltu.se

P. Tretten
e-mail: phillip.tretten@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_52

1 Introduction

The consequences of data quality are significant to businesses, governments and society in general [1]. Data is considered as one of the most important asset for organizations. Its strategic value leads to reconsider the importance of maintaining adequate levels of quality in data that is managed and used by applications, especially in Web Applications as the main organizational showcase. Data quality is a key component of the quality of information and most business processes depend on the quality of data. The existence of poor quality data contributes to unsatisfactory information, unusable results and dissatisfied users [2]. Data of poor quality causes customer dissatisfaction, lost revenue and higher costs associated with additional time to reconcile data. Data of poor quality can lead to a loss of credibility in a system and higher risks of noncompliance with regulations. Data of poor quality increases consumer costs, increases taxes, decreases shareholder value and can cause mission failure [1].

It is essential for consumers and decision makers to know that they are using credible and high quality data. Data Quality (DQ) problems are obvious to observers in the information age nowadays. This subject has been studied and cleared in literature in different area i.e. (production, healthcare, maintenance, aviation, business, etc.).

Any DQ improvement plan must begin with the assessment of the affected scenarios to identify the common roots of the detected problems. The assessment involves having values for DQ measures. The main intention of these measures is to provide a quantitative meaning about how much data quality dimensions are achieved in order to enable an adequate management [3].

The knowledge is in the human mind while information is outside it: in people, written texts and cultural elements. For this reason it is important to check if this information complies as much as possible with expected data quality characteristics, helping to reduce or avoid confusion, stress or problems in the human mind and its patterns of logic [2].

The importance of controlling data quality is crucial to improving the processes that cause poor data quality. The best results in monitoring data are obtained when measurements are automatically updated, varying the content of databases under control, and the radar chart is available on line [2]. Although DQ literature counts with a great amount of measurement proposals, it has still a lot of open standardization and unification problems [3]. The quantity of data handled by information systems is increasing worldwide. Particularly in the World Wide Web, the rise of web services, where data change frequently, it is necessary to pay the attention to standards which can help them deal with such complexities [3].

In order to solve this problem, international standard organization has released ISO 8000 standard. This standard is the first from ISO that is dedicated to DQ management and measurement. It is useful to provide verification of data quality such as, in particular, accuracy, completeness, and consistency, in a standardized way. These are aspects of the content value of the information.

Each aspect of DQ can be measured defining an algorithm, a specific method and a level of the target value that it is necessary to achieve, depending on the context of use. One of the important goals of the data quality analysis is to guarantee not only the high level of expected quality for each feature, but also to understand how the organization can support this quality. In this paper, a DQ assessment model is developed based on ISO 8000 standard.

2 Definitions

Data is defined in the standard as “a reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing”. It is a representation of the perception of the real world. Data can be considered the base of information and digital knowledge and takes into account all data types, such as texts, numbers, images, and sounds [3]. Data can be defined also as a symbolic representation of something that depends, in part, on its metadata for its meaning while data set is a logically meaningful grouping of data. Master data is the data held by an organization that describes the entities that are both independent and fundamental for that organization, and that it needs to reference in order to perform its transactions, see Fig. 1 [4].

Information knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts that within a certain context have a particular meaning. Information system is one or more computer systems and communication systems, together with associated organizational resources such as human, technical, and financial resources that provide and distribute information [6].

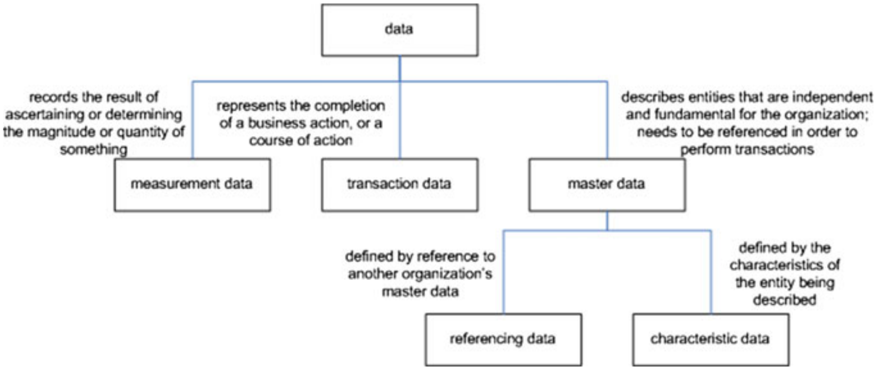

data
records the result of ascertaining or determining the magnitude or quantity of something
represents the completion of a business action, or a course of action
describes entities that are independent and fundamental for the organization; needs to be referenced in order to perform transactions
measurement data
transaction data
master data
defined by reference to another organization's master data
defined by the characteristics of the entity being described
referencing data
characteristic data

Fig. 1 Taxonomy of data (for master data) [5]

Quality is the degree to which a set of inherent characteristics fulfils requirements. Data quality involves data being fit for use by consumers. According to ISO 8000, data quality involves the following principles [4]:

a. Data being fit for purpose; i.e., the decision it is used in.
b. Having the right data, in the right place, at the right time.
c. Meeting agreed customer data requirements.
d. Preventing the recurrence of data defects by improving processes to prevent repetition and eliminate waste.

The requirements are the needs or expectations that are stated, generally implied or obligatory. Data quality management is the coordinated activities to direct and control an organization with regard to data quality. ISO 8000 focus is to provide management, control and measurement of the following data quality aspects: provenance, accuracy and completeness [4].

Data provenance record is a record of the ultimate derivation and passage of a piece of data through its various owners or custodians. Data accuracy is closeness of agreement between a property value and the true value. While data completeness is the quality of having all data that existed in the possession of the sender at time the data message was created [4].

3 Data Quality Dimensions Based on ISO 25012

The data quality dimensions according to International Standard categorizes quality attributes into fifteen characteristics considered by two points of view: inherent and system dependent.

Inherent data quality Inherent data quality refers to the degree to which quality characteristics of data have the intrinsic potential to satisfy stated and implied needs when data is used under specified conditions. From the inherent point of view, data quality refers to data itself, in particular to:

- Data domain values and possible restrictions
- Relationships of data values (e.g. consistency);
- Metadata.

System dependent data quality refers to the degree to which data quality is reached and preserved within a computer system when data is used under specified conditions. From this point of view data quality depends on the technological domain in which data are used; it is achieved by the capabilities of computer systems' components such as: hardware devices (e.g. to make data available or to obtain the required precision), computer system software. In Table 1 below, a summary of DQ dimensions categorizing them into their origins is presented.

Table 1 Data quality model characteristics [6]

Characteristics	Data quality	
	Inherent	System dependent
Accuracy	X	
Completeness	X	
Consistency	X	
Credibility	X	
Currentness	X	
Accessibility	X	X
Compliance	X	X
Confidentiality	X	X
Efficiency	X	X
Precision	X	X
Traceability	X	X
Understandability	X	X
Availability		X
Portability		X
Recoverability		X

4 Data Quality Measurement

Measurement is ascertaining or determining the magnitude or quantity of something [4]. Data quality measure is a variable to which a value is assigned as the result of measurement of a data quality characteristic [6]. Data quality is dependent both on the quality of the data capture process and the processes used to store, maintain, transfer and present data (ISO/TS, 2009a). De Vaux and Hand indicate that 60–95 % of the total effort in data analysis work is spent on data cleaning [7].

Product quality is managed through quality measurements, reliability engineering, and statistical quality control [8]. Since data is considered as a product, measurement is necessary for its quality control. In addition, setting up a time interval to measure data quality is necessary because certain data lose their importance as time goes by. Although it is desirable to measure data quality without delay after the process of data processing, the measurement time can be adjusted in accordance with characteristics of business tasks [9].

Data quality measurement shall consist of the following activities [9]:

- Data quality measurement: the activity that measures target data in accordance with criteria by tools or manually. For repeated data, measurement can be done by tools. Yet for complicated data, measurement can be done by an expert's judgment.
- Statistical treatment of measured data: the statistical analysis of data quality measurements to support the analysis of the causes of defects and non-conformances.

Data errors are typically discovered by chance in the process of data processing, and the data errors tend to be corrected within user's capability or business scope. Therefore, if data error correction is carried out depending on users only, the number of unidentified data errors will gradually increase. For this reason, it is necessary to measure and inspect data errors continuously and systematically. This process can be performed with SQL programs or data quality profiling tools by data operators [9]. A radar chart can be useful to manage data entities, attributes and characteristics to be analysed, measuring the distance between the actual value and the target value estimated, for instance see Fig. (2). The importance of controlling data quality is crucial to improving the processes that cause poor data quality. The best results in monitoring data are obtained when measurements are automatically updated, varying the content of databases under control, and the radar chart is available on line [2].

This chart (Fig. 2) shows the definition of every DQ dimension described in standard ISO/IEC 25012. In this example, by observing the distance between the maximum and minimum level of coverage, the completeness of data appears very limited.

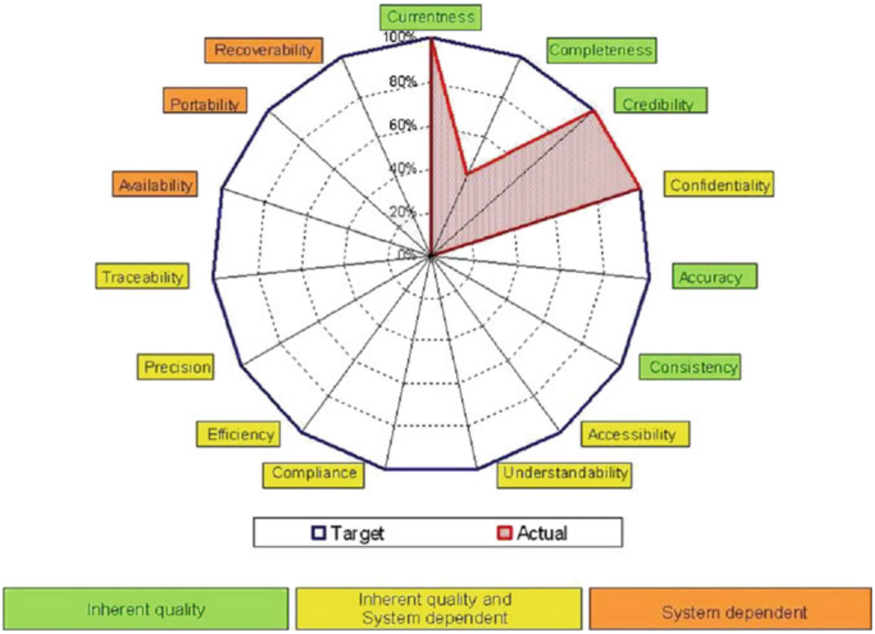


Fig. 2 Data quality chart based on ISO 25012 [2]

5 ISO 8000 Standard

Today, in an increasingly interconnected world, interoperability is more important than ever, and interoperability problems are very costly. Studies of the US automobile sector, for example, estimate that insufficient interoperability in the supply chain adds at least $1 billion in additional operating costs, of which 86 % is attributable to data exchange problems. The adoption of standards to improve interoperability in the automotive, aerospace, shipbuilding and other sectors could save billions [10]. Standardization is the way to achieve interoperability.

There exist a number of quality-related standards developed by ISO/TC. ISO quality models including ISO 9126 and ISO 25010 can be used to support specification and evaluation of software from different perspectives by those associated with acquisition, requirements, development, use, evaluation, support, maintenance, quality assurance and audit of software. The ISO 9126-1 standard distinguishes three different viewpoints for software product quality, internal quality, external quality, and quality in use. ISO 25010 combines internal and external quality models as product quality [11].

General requirements for the quality management to creating process of each product are given in the ISO 9000 and ISO 9001 standards. These standards are mostly process oriented and are intended previously for developers. The ISO 9000 quality management standards are focused on product quality general. ISO 8000 addresses data quality. ISO 8000 specifies fundamental principles of DQ management, and requirements for implementation, data exchange and provenance. ISO 8000 is concerned with [4]:

- the principles of data quality;
- the characteristics of data that determine its quality;
- the processes to ensure data quality.

An organization that implements this part of ISO 8000 shall perform the following actions:

- Perform processes for data quality management that include at least data processing, data quality measurement and correction, data schema design, measurement criteria setup, error cause analysis, data quality planning and data architecture/stewardship/flow management;
- Assign roles for data quality management within their organization;
- Embed processes for data quality management within the organizations business processes.

ISO 8000 is organized as a series of parts, each published separately. The parts of ISO 8000 are organized into the following series [12]:

- parts 1–99: General data quality;
- parts 100–199: Master data quality;
- parts 200–299: Transaction data quality;
- parts 300–399: Product data quality.

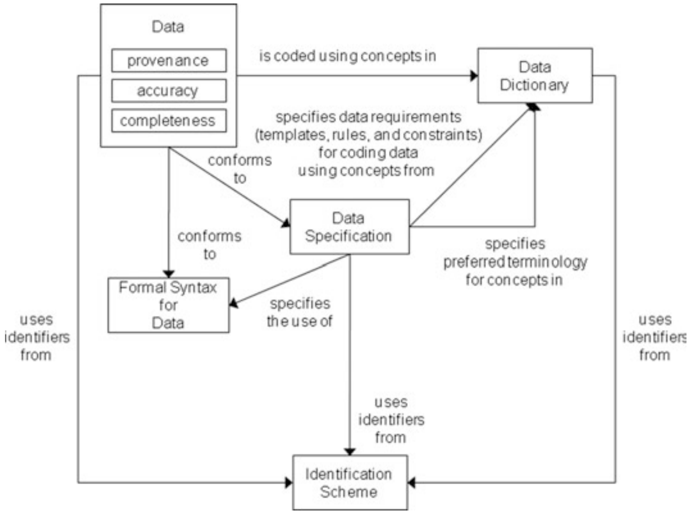


Fig. 3 Graphical depiction of the data architecture [12]

Each of the above series addresses communication within an organization and between two or more organizations.

The data architecture according to the scope of the parts of ISO 8000 is shown in Fig. 3. It is explained in the standard as the following:

- Data includes information about data provenance, data accuracy, and data completeness.
- Data is coded using concepts in a data dictionary.
- Data conforms to a data specification.
- Data conforms to a formal syntax.
- A data specification specifies data requirements for coding data using concepts from a data dictionary.
- A data specification specifies preferred terminology for concepts in a data dictionary.
- A data specification specifies the use of a formal syntax.
- Data, data specifications, and data dictionaries use identifiers from an identification scheme.

6 Data Quality Measurement Model

ISO 8000 includes terms relating to DQ. DQ dimensions included in the standard are completeness, accuracy and provenance. These terms can be listed as follows:

- Data quality management: coordinated activities to direct and control an organization with regard to DQ.

- Data provenance record: record of the ultimate derivation and passage of a piece of data through its various owners or custodians
- Data accuracy: closeness of agreement between a property value and the true value
- True value: value that characterizes a characteristic perfectly defined in the conditions that exist when the characteristic is considered
- Accepted reference value: value that serves as an agreed-upon reference for comparison
- Authoritative data source: owner of a process that creates data
- Data completeness: quality of having all data that existed in the possession of the sender at time the data message was created

The UML class diagram for the high-level data model is given in Fig. 4.

A data_dictionary is a collection of data_dictionary_entry objects that allows lookup by entity identifier. A data_dictionary_entry is a description of an entity containing, at a minimum, an unambiguous identifier, a term, and a definition. A data_record is a data_object that is a set of property_value_assignment objects. A data_set is a data_object that is a set of data_record objects, which may be ordered

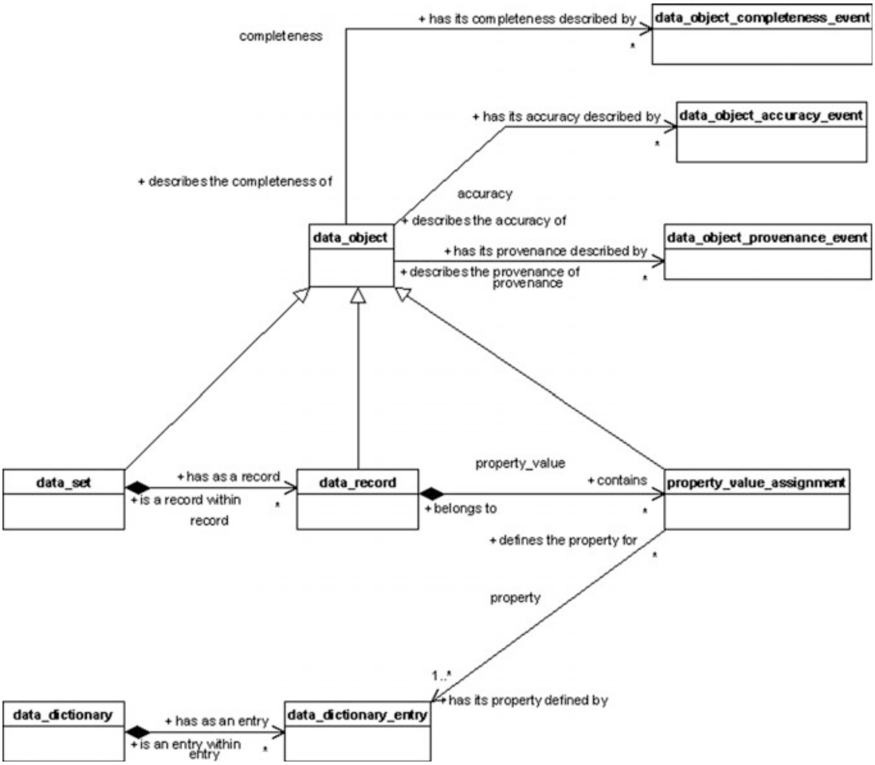

+ has its completeness described by
data_object_completeness_event
completeness
*
+ has its accuracy described by
data_object_accuracy_event
*
+ describes the completeness of
accuracy
+ describes the accuracy of
data_object
+ has its provenance described by
data_object_provenance_event
+ describes the provenance of
provenance
*
data_set
+ has as a record
+ is a record within
record
*
data_record
property_value
+ contains
property_value_assignment
+ belongs to
*
+ defines the property for
*
property
1..*
+ has its property defined by
data_dictionary
+ has as an entry
+ is an entry within
entry
*
data_dictionary_entry

Fig. 4 UML class diagram for high-level data model [5]

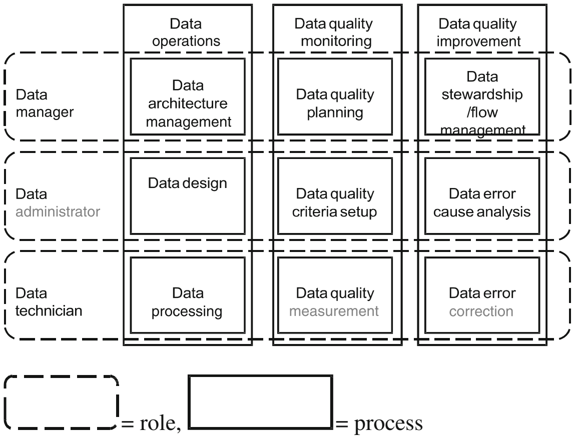


Fig. 5 Data quality management framework (ISO/TS 2011a)

or partially ordered. A data_object is anything that is used to signify something else. A data_object_accuracy_event is an event for which data accuracy information is recorded. A data_object_completeness_event is an event for which data completeness information is recorded. A data_object_provenance_event is an event for which data provenance information is recorded. A property_value_assignment is a data_object that is a pair of a value and an identifier to a property defined in a data dictionary (Fig. 5).

The data operations process identifies factors that affect data quality and ensures data is available at the right place in a timely manner. This top-level process shall consist of the following processes:

- Data architecture management; the process that manages organization-wide data architecture from the integrated perspective to use data in distributed information systems with consistency and therefore ensure data quality.
- Data design; the process that designs data schema, and implements a database to make data users apply data without mistake and ensure data quality.
- Data processing; the process that creates, searches, updates, deletes data in accordance with guidelines of data operations.

The data quality improvement process corrects data errors detected and eliminates root causes of the data errors by tracing and identifying them. In order to support the top-level process effectively, adjustment of data stewardship in accordance with data flows tracing is required. This process has the function of process improvement not only data quality improvement. Processes for data management are improved at the data administrator level while business processes at the data manager level. This top-level process shall consist of the following processes:

- Data stewardship and flow management; the process that analyses data operations and data flows among businesses or organizations, identifies responsible parties and their data operation systems which influence data quality, and manages the stewardship of data operations.

- Data error cause analysis; the process that analyses root causes of data errors and prevents a recurrence of the same errors fundamentally.
- Data error correction; the process that corrects the data that turns out erroneous.

The three roles in the framework are responsible for performing the processes in the framework. These roles shall be: data manager; data administrator; data technician. The data manager shall perform the following processes within the framework:

- Data architecture management.
- Data quality planning.
- Data stewardship and flow management.

The data manager performs the role that directs a guideline for master data quality management in compliance with objectives of an organization, manages factors that impact data quality at an organization level, and establishes the plans for performing data quality activities in the organization. Along with each major top-level process, the data manager maintains data consistency in individual information systems through the organization-wide data architecture management, and analyzes factors that affect data quality in data quality planning. In addition, the data manager takes a role of granting data administrators the authority to trace and correct data over the information systems or organization.

The data administrator shall perform the following processes within the framework:

- Data design.
- Data quality criteria setup.
- Data error cause analysis.

The data administrator controls and coordinates over data technicians by defining criteria required to maintain the quality of master data, and prevents a recurrence of the same data errors by analyzing the causes of errors or designing data schema. In general, supporting resources and guidelines to data technicians, the data administrator carries the data quality plan into practice to achieve the objectives set by the data manager.

The data technician shall perform the following sub-processes within the framework:

- Data processing.
- Data quality measurement.
- Data error correction.

The data technician creates, reads, modifies, and deletes data as per guidelines of data quality management set by the data administrator, and measures data quality and corrects erroneous data as a result of the measurement. While the data manager or administrator can handle data even across its own business scope in accordance with data flows, the data technician handles data within its business scope.

7 Conclusions and Discussions

The phenomenal global growth of the Internet coupled with the ever-increasing sophistication of online technologies and emergent spread of information require even higher quality data and information. The lack of standardized approaches regarding requirement specification and evaluation lead us to discover this issue in depth. The lack of a common terminology may affect the data quality measurement of the DQ research community. This paper presents a proposal for DQ Measurement to fulfill the ISO 8000 standard, provides a set of the main terms related to data quality measurement. This set has been elaborated by analyzing the data quality measurement provided by the standard.

Data quality is a major concern also in Data analytics projects. One of the first conclusions we raised is that quality-in-use becomes more representative when it comes to measure the level of quality in big datasets composed by several datasets coming from different sources, with probably different formats, and at different velocities. Completeness, accuracy and provenance attributes are provided by ISO 8000 standard. However, extending this standard to include other attributes, such as consistency, is necessary. These dimensions are to be measured according to data quality dimensions defined on each one of the datasets.

We have proposed a framework to specify quality requirements for data for employing a minimalist and standard approach by reusing and extending the ISO 8000 quality models' characteristics. In doing so, we have added information consistency, and value added as new characteristics and have carried out grouping of characteristics based on their conceptual similarities in existing ISO 25012 standard while also combining and integrating characteristics from previous research.

References

1. ISO/TS (2009) Technical specification ISO/TS 8000-120:2009(E)—data quality—Part 120: master data, exchange of characteristic data: provenance. Geneva, Switzerland
2. Natale D (2011) Complexity and data quality. Poster e Atti Conferenza, pp 13–16
3. Caballero I, Verbo E, Calero C, Piattini M (2007) A data quality measurement information model based on ISO/IEC 15939. ICIQ, pp 393–408
4. ISO/TS (2012) International Standard ISO 8000-2:2012(E)—data quality—Part 2: vocabulary. Geneva, Switzerland
5. ISO/TS (2009) Technical specification ISO/TS 8000-100:2009(E)—data quality- Part 100: master data, overview. Geneva, Switzerland
6. ISO/TS (2008) Technical Specification ISO/IEC 2512:2008—software engineering—Software product Quality Requirements and Evaluation (SQuaRE)—Data quality model. Geneva, Switzerland
7. Shirai Y, Nichols W, Kasunic M (2014) Initial evaluation of data quality in a TSP software engineering project data repository. In: Proceedings of the 2014 international conference on software and system process, pp 25–29

8. Wang RY, Kon HB, Madnick SE (1993) Data quality requirements analysis and modeling. In: Proceedings of the ninth international conference on data engineering, pp 670–677
9. ISO/TS (2011a) Technical specification ISO/TS 8000-1:2011(E)—data quality—Part 150: master data: quality management framework. Geneva, Switzerland
10. Folmer E (2011) Quality model for semantic IS standards
11. Rafique I, Lew P, Abbasi MQ, Li Z (2012) Information quality evaluation framework: extending ISO 25012 data quality model. In: Proceedings of world academy of science, engineering and technology, p 65
12. ISO/TS (2011) Technical specification ISO/TS 8000-150:2011(E)—data quality—Part 1: overview. Geneva, Switzerland

Data Quality Assessment of Automatic Wheel Profile Measurement Systems

Matthias Asplund, Stephen M. Famurewa and Matti Rantatalo

Abstract The aim of this paper is to present a method for the quality assessment of data from a condition monitoring system for rolling stock wheels to ascertain if the data have the right quality to be used for further analyses. This quality assessment will also show if there are variations between different measurement units for the same system, and if there are relations between different wheel parameter measurements, speed and time. The assessment of data is accomplished using the quality dimension freedom of error. There are two different data sources, namely an automatic wheel profile measurement system and a manual wheel profile measurement device. The manual measurements of wheel profiles are used for verifying the accuracy of the automatic wheel profile measurements, which constitute the larger data set. The proposed method for evaluating the data quality is demonstrated using the data from a specific condition monitoring system. The results show some inconsistencies indicating that this system lacks quality in the dimension of freedom of error and that there is need for internal calibration or self-adjustment of the studied system for quality reasons.

Keywords Railway · Wheel profile measurement · Maintenance · Data assessment · Condition monitoring

M. Asplund (✉) · S.M. Famurewa · M. Rantatalo
Luleå University of Technology, Luleå, Sweden97187
e-mail: matasp@ltu.se

S.M. Famurewa
e-mail: stefam@ltu.se

M. Rantatalo
e-mail: matti.rantatalo@ltu.se

© Springer International Publishing Switzerland 2016

U. Kumar et al. (eds.), *Current Trends in Reliability, Availability, Maintainability and Safety*, Lecture Notes in Mechanical Engineering,
DOI 10.1007/978-3-319-23597-4_53

1 Introduction

Condition-based maintenance (CBM) employs a preventive maintenance approach that includes one or more of the following actions: condition monitoring, testing, inspection, analysing and the ensuing maintenance actions [1]. It is really important to have a high data quality for the CBM process to detect the state of the item, which is necessary to know if one is to take correct decisions. CBM is used in the maintenance of the track alignment, which is based on data supplied by a condition monitoring system (CMS) installed in recording cars.

Condition monitoring (CM) is an activity that can be performed manually or automatically to measure the characteristics and parameters of the actual state of an item. CM is conducted continuously or at intervals, and usually in the operational state [1]. Data from CMS are an essential tool for improving and optimising maintenance decision making, and make a significant contribution to the boosting of key business drivers [2]. Therefore, such data need to give reliable information as input to the maintenance process; a description of the maintenance process is given in the dependability management standard IEC 60300-3-14 [3]. An applicable and effective condition monitoring task demands the fulfilment of a number of requirements, examples of which are: measurable parameters that can be used to detect the status of the equipment, a reasonably consistent interval between the onset of failure and the detection of failure, and cost-effectiveness [4].

The data quality exerts an influence on the maintenance actions taken. The railway is a one-dimensional system and poor data quality will probably lead to improper maintenance decisions, such as decisions to take wrong actions or no action at all. This will in turn lead to a failure that consumes availability and capacity [5] in an already highly occupied system, and secondary failures can arise. This will entail excessive downtime for the infrastructure and the rolling stock [6].

Considering the inherent characteristics of the capacity of a railway system, the failure-driven capacity-consuming events within a railway network should be kept to a minimum. This can be achieved by the use of an appropriate CBM which can use CMS to detect, diagnose and predict failure events at an early stage to support the maintenance decisions. A review of CMS for railway systems, relating to the condition of the rolling stock and with relevance to both the operator and the infrastructure manager, is presented in [7].

CMS can be either reactive or proactive in their approach, but the focus of CBM requires a proactive monitoring system which gives room for adequate prognostics and a preventive maintenance approach. One characteristic of the reactive approach is that it provides a better possibility of finding and understanding trends in the deterioration of the affected item, and then analysing the condition of the item in question for maintenance decision support [8, 9].

The fact that this article deals with data assessment motivates a brief presentation of this topic. Data can be assessed according to 16 types of data quality dimensions: accessibility, appropriate amount of data, believability, completeness, concise representation, consistent representation, ease of manipulation, freedom of error, interpretability, objectivity, relevancy, reputation, security, timeliness, understandability and added value [10]. There are also other ways of classifying data quality dimensions, one of which uses the following five dimensions: usefulness of content, adequacy of information, usable accessibility, privacy/security and interaction [11]. Lee et al. [12] present a summary of academics' views of information quality, focusing only on subjective data assessment. Pipino et al. [13] provide a method for both subjective and objective data assessment and for presenting the outcome in a matrix of four quadrants.

One often finds general descriptions of the CMS for railway applications, and information is also available concerning railway applications that combine track measurement data [14]. The optimal use of data from condition monitoring systems requires the data to have a high credibility, availability, and accuracy, as well as good repeatability. This paper deals with data from an automatic wheel profile measurement system (WPMS) in service and assesses the data according to the quality dimension of freedom of error.

The paper presents a method for the data quality assessment of automatic WPMS, as well as a case study of automatically generated wheel data and manually measured wheel data. The data are examined from the repeatability point of view by using a method that includes the following elements: data collection, cleaning and visualization, comparison of two measurement units with each other, and comparison of the WPMS data with manually measured data. The aim of this procedure is to ascertain whether this method can be used for this kind of assessment and whether the data can be used for further maintenance optimisation based on the prediction of degradation and for further modelling and simulation purposes. The paper concludes with a discussion and an assessment of the outcome of the data quality assessment and how the data can be used for enhancing the maintenance decision.

Section 2 of the paper starts by presenting condition monitoring and regular wheel wear such as tread and flange wear, and wear related to fatigue, and then describes the automatic and manual wheel measurement that generated the data for this case study. Section 3 deals with the method of data assessment, after which the case study is described in Sect. 4. Then the paper ends with a presentation of conclusions drawn.

2 Condition Monitoring and Wheel Failures

CM is a task designed to detect failures. CMS can decrease the operational risk, enhance the performance and in the long run contribute to cost reduction. CM methods can be categorized into analysis, process monitoring, performance monitoring, functional testing and inspection [15]. The terms monitoring and inspection are well defined in the maintenance standard [1].

2.1 Condition Monitoring Systems

There are different types of CMS for railway rolling stock, and they employ different approaches, are used for different applications, and exploit a large variety of technologies. The different CMS can be divided into way-side and on-board (mobile) monitoring systems, with the former usually being used for the rolling stock and the latter being used for the infrastructure. However, there are also way-side monitoring systems for the infrastructure, such as camera-based monitoring systems for switches and crossings [16]. This section briefly describes the CMS that are used in the railway system.

2.1.1 Reactive and Proactive CMS

Some examples of reactive systems which are installed to limit damage are dragging equipment detectors, hot box detectors, hot/cold wheel detectors and sliding wheel detectors. Although this category of equipment is useful, it has the limitation of not being able to capture potential failures early enough before failure to allow proactive decision making. Some examples of proactive systems are vehicle inspections, hunting vehicle and bogie performance monitoring, wheel condition monitoring and acoustic bearing detectors. Proactive systems provide a better possibility of understanding trends in the deterioration of vehicle components and analysing the condition of the affected item for maintenance decision support [8, 9].

2.1.2 Way-Side Condition Monitoring

In the Swedish railway network there are almost 200 way-side CMS. The purpose of these systems is usually to monitor the interface between the infrastructure and the rolling-stock, for example to detect high wheel forces and hot boxes, and to perform measurements on the pantograph. Most of these systems adopt a reactive approach. Many descriptions of way-side monitoring systems and many assessments of their potential are to be found in the literature [9]. The only proactive way-side monitoring system installed in the Swedish railway network is the automatic wheel profile measurement system.

2.1.3 On-Board Condition Monitoring

The on-board CMS employ a more proactive approach than the way-side CMS, and one example of the former is the track quality monitoring system. The frequency of the monitoring depends on the track classification. Usually a track section is subjected to monitoring one to six times each year, and when a defined threshold is reached, maintenance action is ordered [17]. Condition monitoring of the infrastructure is described in [6].

2.2 *Wheel Wear*

There are two main types of wheel wear, regular wear and irregular wear [18]. One example of irregular wear is out-of-round wheels, with failure modes such as flats and eccentric wheels. Regular wheel wear can be monitored through profile measurements. This section treats only tread and flange wear and rolling contact fatigue, which are examples of regular wheel wear. Different traffic types result in different dominant types of wheel wear; in regular traffic the major reason for re-profiling is flange and tread wear, while in heavy haul traffic the dominant reason for re-profiling is rolling contact fatigue (RCF) [19].

2.2.1 Tread and Flange Wear

The wheel profile tends to go from a conical shape to a more concave shape when it wears [20]. Figure 1 shows an example of a worn wheel profile together with an original wheel profile shape, with the wheel measurements that explain the condition of the wheel: the flange height (Sh), flange width (Sd), flange slope (qR) and tread hollowing (TH). The major reasons for re-profiling wheels (excluding wear caused by heavy haul traffic) are wear of the flange and hollow wear.

A wheel profile is not constant around the whole wheel; there is an average roundness variation of the Sh of 0.131 mm, while the average variation of the Sd is 0.145 mm and that of the TH is 0.087 mm [21].

2.2.2 RCF

The reason for RCF is in many cases bad wheel-rail contact, and the wheel-rail contact is dependent on the wheel and the rail profiles. The high axle load (30

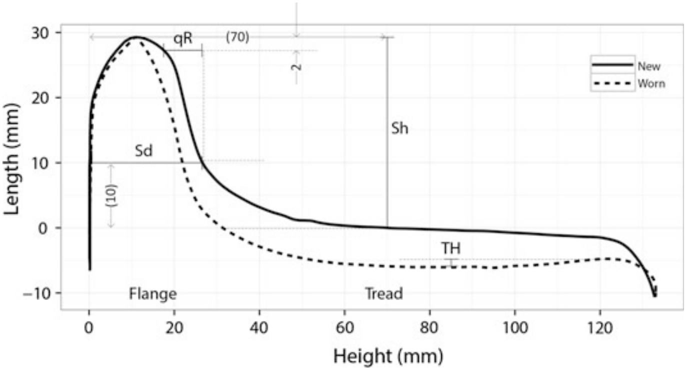


Fig. 1 Original and worn wheel profiles, with the wheel profile parameters illustrated in the figure. *Sh* flange height, *Sd* flange thickness, *qR* flange slope, and *TH* tread hollowing

metric tons) of the iron ore trains on the Iron Ore Line causes a large amount of RCF failures of the wheels and rail. There are many on-going activities being performed to reduce RCF, for instance investigations of locomotive wheel failures [22] and wheel data analyses [19, 23].

For the wagon wheels, RCF is a less serious problem. The locomotive wheels need to be re-profiled between 10,000 and 20,000 km, while the wagon wheels can survive for up to 200,000 km before re-profiling is needed. Wheels have different RCF patterns, and the pattern zones can be divided into three different types of RCF zones, RCF 1, RCF 2 and RCF 3. RCF 1–3 are caused by different phenomena and their crack structures differ [24].

3 Automatic Wheel Profile Measurement

The WPMS can measure the wheel profile for speeds up to 140 km/h. The advantages of the automatic WPMS are improved safety, reliability and efficiency, as well as more effective maintenance actions leading to a higher train availability and less failure-driven capacity consumption [9].

The only WPMS installed in Sweden is located in the track section which lies in the southernmost part of the Iron Ore Line. This system has been in use since autumn 2011. The data are delivered to the operator for improvement of the wheel maintenance, and in the future will also be delivered to the infrastructure manager for development of the track maintenance [25].

The wheel profile measurement equipment consists of four sub-units, two on the gauge sides and two on the field sides of the rails. These units contain a laser, a high-speed camera, and an electronic control system. When a train passes the boxes housing these units, the first wheel triggers a sensor and the protective cover opens, the laser beam starts to shine, and then the camera takes pictures of the laser beam projected onto the surface of the passing wheels. These pictures are converted to wheel profiles and the system describes the flange height, flange width, flange slope, tread hollow and rim thickness. The measurement accuracy for this camera- and laser-based technology is around 0.1 mm [7]. The WPMS consists of two different main units, the far and the near unit, see Fig. 2.

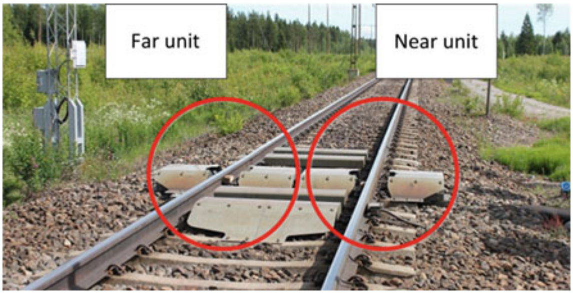


Fig. 2 The automatic WPMS with its near and far unit

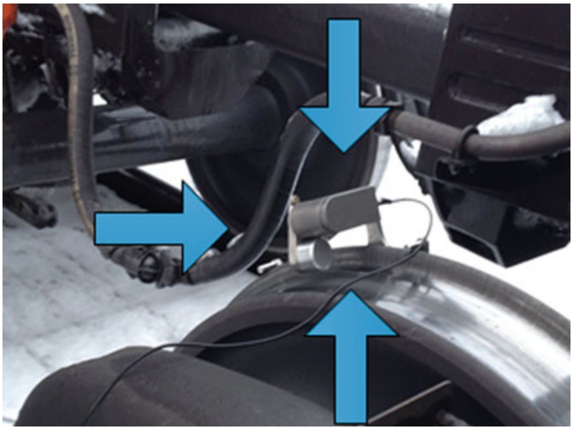


Fig. 3 Manual measurements performed with the hand-held MiniProf measurement equipment on an iron ore wagon wheel

3.1 Manual Wheel Profile Measurement

An established way to measure the wheel profiles manually is to use hand-held portable measurement equipment. In the market, different types of manual measurement equipment are available. Compared with the automatic WPMS, this measurement method is really time-consuming; for example, measuring a wagon with eight wheels takes around 5–10 min. In this study the manual measurements were performed with MiniProf measurement equipment [26]. This manual measurement equipment is attached to a wheel with a magnetic foot, and the measuring arm is moved manually, see Fig. 3. The data are stored in a hand-held unit. The measurement accuracy for the MiniProf is ±0.9 µm [27].

4 Method of Data Assessment

The method of data assessment is presented in this section. The goal is to verify the data quality dimension freedom of error, which can be defined as meaning the extent to which data are correct and reliable [10]. The data assessment process is presented in Fig. 4.

The first step is to collect data for analyses, and the data can come from one or many sources. The literature from this field confirms that the parameters that are to be studied must have the right characteristics for condition monitoring purposes. It is necessary to have a data set which represents the process in question and which possesses the right data size.

The second step is data cleaning, which is performed manually by removing unrealistic data from the data set and checking it for missing data. Depending on the size of the data, advanced cleaning techniques can be adopted, for instance the definition of outliers.

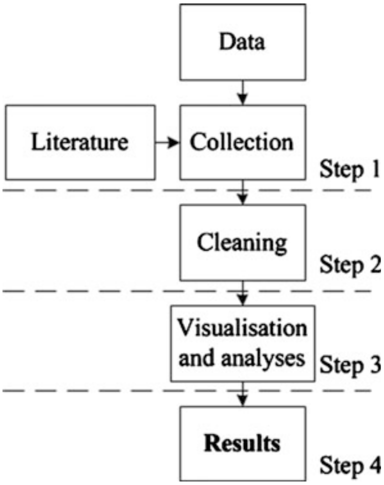


Fig. 4 The method of data assessment

The third step is the data visualisation and analyses. This starts with a goodness-of-fit test to determine the statistical distribution that adequately models the randomness of the data. Then the visualisation and analyses of the data follow. This is accomplished in two steps, first with all the data and then with a small part of the data, in order to ascertain the influence of the data size. A freedom of error test is conducted by comparing the data with data from the same but parallel processes and with reference data. The analysis includes a goodness-of-fit test, a graphical test, a paired T-test for comparison between different measurement techniques, and regression tests of key parameters.

The fourth step is to show the results from the data cleaning and data analyses.

5 Case Study of Measurement Data from a WPMS

This section presents the following data analyses performed within a case study: a paired T-test of data from the far and near units, a comparison between automatic and manual measurements, and a regression analysis, according to the proposed data assessment method.

5.1 Data Collection

The data collected concern the wheel parameters Sd, Sh, qR and TH, which are defined in Fig. 1. The data come from all eight wheels of iron ore wagon 4907. The data compilation was performed using the Excel software. The number of automatically generated data was 80, for eight wheels on the same wagon, and the number of manual measurements was 16, i.e. two samples for each wheel. The automatically generated measurements were taken at random locations on the wheel circumference. The manual measurements were taken at two places on the circumference of the wheel, 90° from each other. The automatically generated data come from trains operating between 13 and 22 November 2014. The manual measurements for each wheel were conducted on 11 November 2014. Figure 3 shows the manual measurement of a wagon wheel. The distance covered by the wagon providing the data samples, wagon 4907, between 11 and 22 November was 3068 km, as estimated from the planned operational profile of the wagon. This distance is only 1.5 % of the distance between two consecutive wheel re-profiling actions, and consequently the wheel wear occurring over this short distance can be ignored.

Figure 5 shows the iron ore wagon relative to the position of the WPMS. Furthermore, the figure shows the position of wheels 1–8.

5.2 Data Cleaning

After the manual data cleaning, 73 out of a total of 80 measurements from the WPMS and from wagon 4907 could be used; seven of the measurements failed, and the missing data concern measurements performed on 13, 14, 19, 21 and 22 November. All the missing data concern measurements made by the far unit. All of the 16 manual measurements could be used.

5.3 Data Visualisation and Analyses

The velocity of the measured trains on the different measurement occasions was between 50 and 72 km/h, and the mean velocity of all the passages is 63.1 km/h.

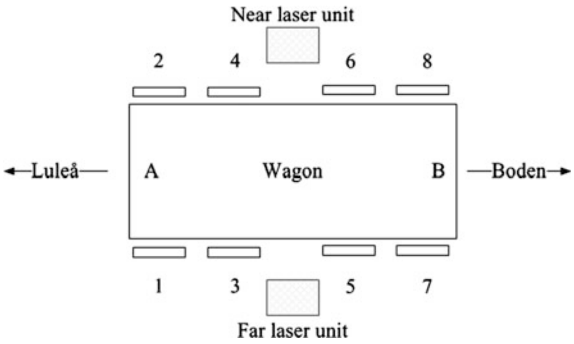


Fig. 5 A view from above over an iron ore wagon at the position of the WPMS; the wheels are numbered from 1 to 8 and the WPMS has a near laser unit and a far laser unit

The Anderson-Darling test shows that the Sh and qR values from the far and near unit, and the Sd values from the near unit can be suitably modelled with a normal distribution. Furthermore, the Sd values for the far unit have a Weibull distribution.

5.3.1 Data Visualisation

The data visualisation starts with the descriptive statistics of all the data. The standard deviation for the data for the wheel parameters flange thickness (Sd), flange height (Sh) and flange angle (qR), defined in Fig. 1, is presented in Fig. 6, for all eight wheels and for the far and near units. The number of samples for the far unit is 33 and that for the near unit is 40. The largest standard deviation is 2.355 mm and this concerns the Sd for wheel 1 as measured by the far unit. The standard deviation for the Sd for all the wheels is 1.4937 mm. The Sh has a smaller standard variation, with a maximum of 0.370 mm, and the standard deviation of Sh is 1.9067 mm for all the wheels. The largest standard deviation for qR concerns wheel 2 and is 0.559, and the standard deviation of all the wheels for qR is 1.0322 mm. Furthermore, Fig. 6 shows that the far unit has a larger standard deviation than the near unit.

The mean values for all the wheels for Sd, Sh and qR are 28.2, 32.00, and 10.53 mm, respectively. The histograms and density plots for the near and far units, respectively, and for the Sd, Sh and qR measurements, respectively, are shown in Figs. 7, 8, and 9; for the far unit 33 measurements were used and for the near unit

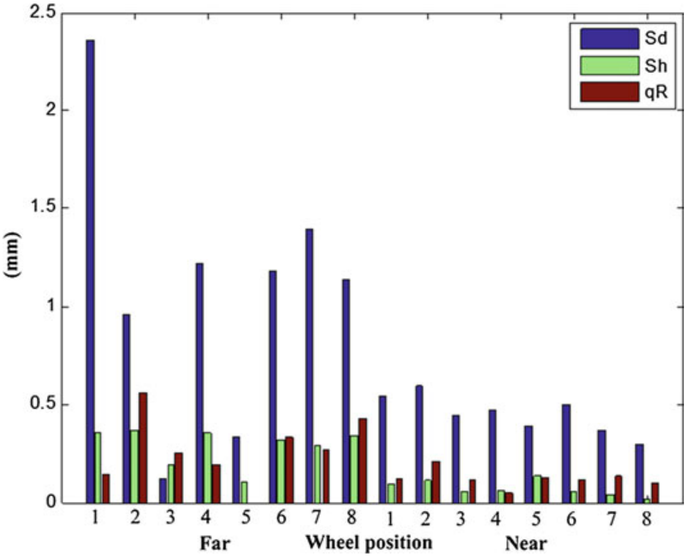


Fig. 6 The standard deviation for the wheel parameters Sd, Sh and qR for the far and near unit

40 measurements. The density differs between the near and the far unit in that the near unit has a weight to the left and the far unit has more weight to the right, and the bars differ too between the far and the near unit in Fig. 7. The mean value for the near unit is 28.00 mm and the standard deviation is 1.390 mm. The mean value and the standard deviation for the far side are 28.32 and 1.655, respectively.

The Sh also differs between the far and the near unit in that the far unit has more weight to the left-hand side and the near unit has more weight to the right-hand side, see Fig. 8. One can also observe in the figure that the bars differ between the far and the near unit. The mean value for the near unit is 32.38 mm and the standard deviation is 1.846 mm. The mean value and the standard deviation for the far side are 31.36 and 1.929, respectively.

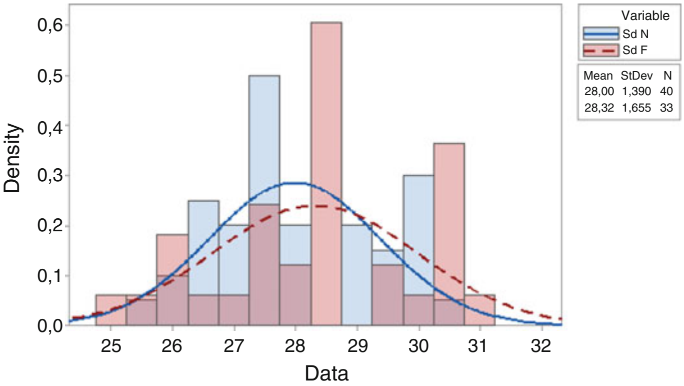


Fig. 7 Histogram and density plot of the flange thickness (Sd) for the near and far unit, with the mean measurement differing by 0.32 mm

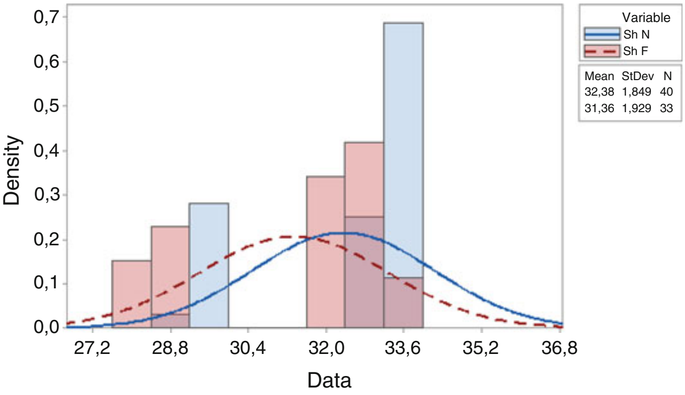


Fig. 8 Histogram and density plot of the flange height (Sh) for the near and far unit, with the mean measurement differing by 1.02 mm

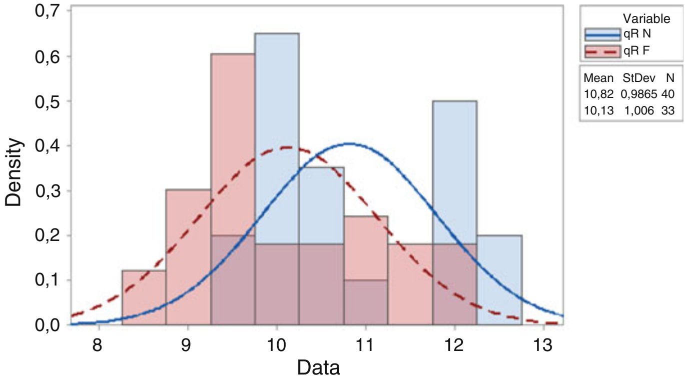


Fig. 9 Histogram and density plot of the flange angle (qR) for the near and far unit, with the mean measurement differing by 0.69 mm

The qR also differs between the far and the near unit in that the far unit has more weight to the left-hand side and the near unit has more weight to the right-hand side, see Fig. 9. One can also observe in the figure that the bars differ between the far and the near unit. The mean value for the near unit is 10.82 mm and the standard deviation is 0.9865 mm. The mean value and the standard deviation for the far side are 10.13 and 1.006, respectively.

If the far and the near unit had had the same measurement performance, the histograms for each unit in Figs. 7, 8 and 9 would have been equal, but they differ according to the figures and the values. A comparison between the far and the near unit shows a larger standard deviation for all the measurements of the far unit.

The next step is a visualisation using data from only one unit and one wheel. Figure 10 shows histograms of six different auto-generated measurements from the near unit for wheel 5.

For the Sd, the measurements range between 25.6 and 26.6 mm (a range of 1.0 mm), and the distribution is asymmetrical, with more weight on the higher side. Furthermore, the Sh varies between 33.2 and 33.5 mm (a range of 0.3 mm) and the shape of the distribution is even. The qR shows a range between 10.1 and 10.4 mm (a range of 0.3 mm), and the TH distribution is asymmetrical, with more weight on the lower values. The standard deviations of these six auto-generated measurements from the WPMS are as follows: $\mathrm{StD}_{\mathrm{Sd}} = 0.3889$, $\mathrm{StD}_{\mathrm{Sh}} = 0.1322$ and $\mathrm{StD}_{\mathrm{qR}} = 0.1265$.

5.3.2 Paired T-Test of Wheel Parameters

A paired T-test of the values of the wheel parameters Sd, Sh and qR obtained from the WPMS will show if there is a difference between the means of the respective

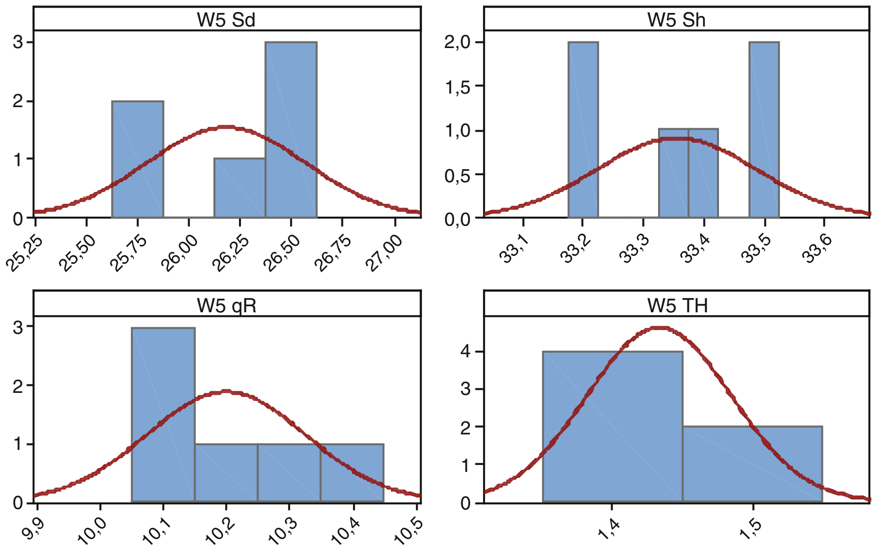


Fig. 10 Histograms of the Sd, Sh, qR and TH measurements from the near unit for wheel 5

parameters measured by the far and near units of the system. The null hypothesis H_0 is according to Eq. 1.

$$\mu_{F-unit} - \mu_{N-unit} = 0. \tag{1}$$

The question is whether H_0 can be rejected. If H_0 is rejected, then one can say that there is a difference between the measurement outputs of the far and near unit. This test was conducted with 28 wheel measurements made by each measurement unit on the same wagon with the level of confidence set to 95 %.

The paired formulas used for these four calculations are as follows: paired T for Sd F–Sd N, paired T for Sh F–Sh N, paired T for qR F–qR N and paired T for TH F–TH N.

Tables 1, 2, 3, and 4 show the mean value, standard deviation and SE-mean (standard error of the mean value) of the far and near measurement unit for the parameters Sd, Sh, qR and TH.

A 95 % confidence interval (CI) was used for the mean difference of Sd: (−0.470; 0.587). The T-test of the mean difference of Sd = 0 (vs ≠ 0) gave the following values: T-value = 0.23 and P-value = 0.823.

Table 1 Parameter results from the T-test for Sd and for the far and near unit of the WPMS

Parameter	Mean	StD	SE mean
Sd F	28.214	1.738	0.328
Sd N	28.156	1.408	0.266
Difference	0.058	1.363	0.258

Table 2 Parameter results from the T-test for Sh and for the far and near unit of the WPMS

Parameter	Mean	StD	SE mean
Sh F	31.460	1.912	0.361
Sh N	32.173	1.900	0.359
Difference	−0.7125	0.3482	0.0658

Table 3 Parameter results from the T-test for qR and for the far and near unit of the WPMS

Parameter	Mean	StD	SE mean
qR F	10.193	1.011	0.191
qR N	10.875	1.054	0.199
Difference	−0.6821	0.3654	0.0690

Table 4 Parameter results from the T-test for TH and for the far and near unit of the WPMS

Parameter	Mean	StD	SE mean
TH F	0.446	0.605	0.114
TH N	0.343	0.527	0.100
Difference	0.1036	0.1290	0.0244

The H_0 cannot be rejected when the P-value is >0.05 and the interval of the mean difference with a 95 % CI covers zero. In other words, the measurements of Sd for the far and near unit do not differ for a CI of 95 %. Figure 11 shows the histogram of Sd_{Far}, $-Sd_{Near}$ with a 95 % CI, and the hypothesis H_0 and $\bar{x}$.

A 95 % CI was used for the mean difference of Sh: (−0.8475; −0.5775). The T-test of the mean difference of Sh = 0 (vs ≠ 0) gave the following values: T-value = −10.83 and P-value = 0.000.

The H_0 can be rejected when the P-value is <0.05 and the interval for the mean differences covers zero. In other words, for a CI of 95 % the measurements of Sd for

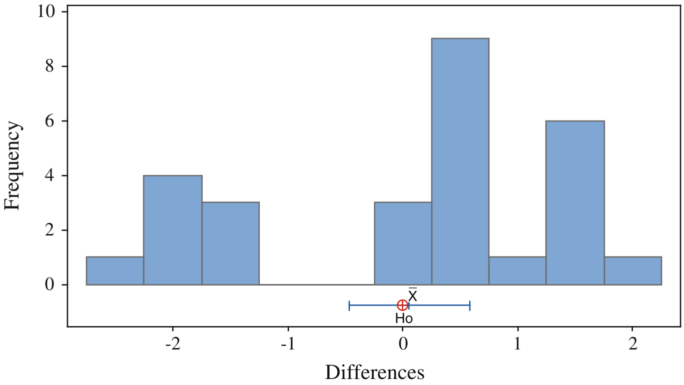


Fig. 11 Histogram of differences with a 95 % CI for the wheel parameter Sd, comparing the far and near unit

the far and near unit will differ. Figure 12 shows the histogram of $\mathrm{Sh}_{\mathrm{Near}}$ -$\mathrm{Sh}_{\mathrm{Far}}$, with a 95 % CI, and the hypothesis H_0 and $\bar{x}$.

A 95 % CI was used for the mean difference of qR: (−0.8238; −0.5405). The T-test of the mean difference of qR = 0 (vs ≠ 0) gave the following values: T-value = -9.88 and P-value = 0.000.

The H_0 can be rejected when the P-value is <0.05 and the interval for the mean differences covers zero. In other words, for a CI of 95 % the measurements of Sd for the far and near unit will differ. Figure 13 shows the histogram and the boxplot of $\mathrm{Sh}_{\mathrm{Near}}$ -$\mathrm{Sh}_{\mathrm{Far}}$, with a CI of 95 %, and the hypothesis H_0 and $\bar{x}$. (Figure 14)

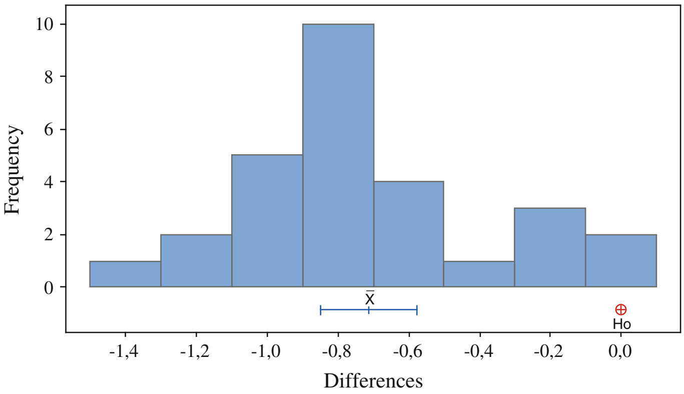


Fig. 12 Histogram of differences with a 95 % CI for the wheel parameter Sh, comparing the far and near unit

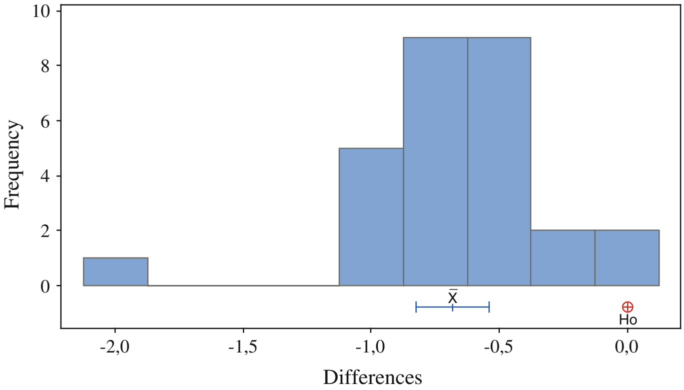


Fig. 13 Histogram of differences with a 95 % CI for the wheel parameter qR, comparing the far and near unit

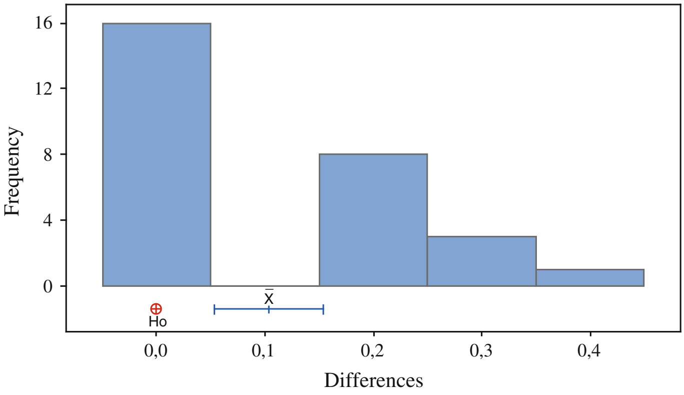


Fig. 14 Histogram of differences with a 95 % CI for the wheel parameter TH, comparing the far and near unit

A 95 % CI was used for the mean difference of TH: (0.0535; 0.1536). The T-test of the mean difference of TH = 0 (vs ≠ 0) gave the following values: T-value = 4.25 and P-value = 0.000.

The H_0 can be rejected when the P-value is <0.05 and the interval for the mean differences covers zero. In other words, for a CI of 95 % the measurements of TH for the far and near unit will differ. Figure 13 shows the histogram of TH_{Near} -TH_{Far}, with a CI of 95 %, and the hypothesis H_0 and $\bar{x}$.

5.3.3 WPMS Profiles Versus Manually Measured Profiles

This comparison between the WPMS profiles and the manually measured profiles can be accomplished since the accuracy of the MiniProf hand-held measurement equipment is higher than that of the WPMS; we assume that $\varepsilon_{MiniProf} \ll \varepsilon_{WPMS}$.

The error is based on Eq. 2:

$$\begin{aligned} \varepsilon_{ij} &= x_{ij}(WPMS) - x_{ij}(Miniprof) \\ i &= 1, \ldots, n \\ j &= 1, \ldots, m \end{aligned} \tag{2}$$

Here x = Sh, Sd, qR and TH, i = the number of wheels, n = 1, and the number of measured parameters m = 4.

The comparison between the manual measurements and the WPMS measurements for wheel 5 (15 L) is shown in Fig. 15. The figure shows two manual measurements together with one measurement from the WPMS. The conformity between the manual measurements is good and there are no crucial differences, but the conformity between the manual and the WPMS measurements is not so good

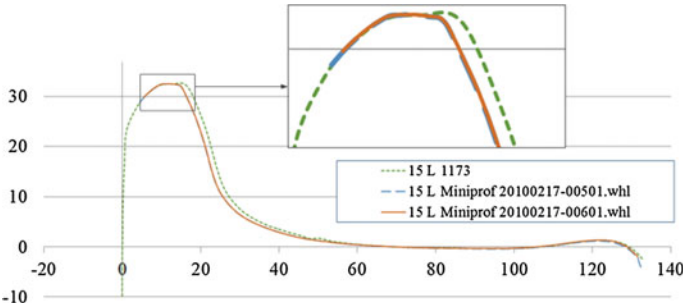


Fig. 15 Comparison between two manual measurements and one WPMS measurement

for the top of the flange and the flange slope, see Fig. 15. The WPMS profile shown in Fig. 15 comes from the far measurement unit.

According to the measurements from the WPMS, the flange top is higher than it is according to the manual measurements.

The wheel measurements for Sd, Sh, qR and TH for wheel 5 are shown in Table 5, and the largest difference between the WPMS and manual measurements concern the Sd, which conforms with Fig. 15 above, with the largest deviation on the flange side.

5.3.4 Regression Analysis

To determine whether there is any relation between selected parameters, regression analyses were performed, using linear regression and analysing only two parameters, β_0 and β_0, at a time, according to Eq. 3.

$$y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i \, i = 1, \ldots, n. \tag{3}$$

Equation 3 expresses the real relationship between the parameters and here the error factor is included; the value of n is 73. By removing the error factor, the mean value of y is obtained, i.e. the $\hat{y}$, see Eq. 4.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \times x \tag{4}$$

Table 5 A wheel measurement from the WPMS compared with a manual measurement for wheel 5 in wagon 4907

Date	Measurement id.	Sd	Sh	qR	TH
141122	WPMS 164135	27.31	32.70	9.60	1.60
141117	MiniProf 20100217-00601.whl	25.70	32.55	9.66	1.68
	Diff.	1.61	0.15	−0.06	−0.08
	In %	6.5	0.5	−0.06	−4.8

S = standard deviation of the y value.

R^2 = coefficient of determination. $0 \leq R^2 \leq 1$.

The R^2 shows how well the model represents the data and gives an indication of the goodness of fit.

The influence of the far and near unit is presented in Table 6, and qR and TH show the largest difference in percent between the units' measurements, i.e. $\boldsymbol{\beta_1/\beta_0}$. The results need to be considered in light of the fact that the value of R^2 is low, which means that the proposed model (Eq. 4) does not fit the values well.

The influence of the train speed between 50 and 72 km/h on the measurement is shown in Table 7. The R^2 value is low and this means that the proposed model (Eq. 4) does not fit these values well. Quite apart from the low correlation with the model, there is no influence of the train speed related to the wheel parameters for the units. The Sd F has a slightly larger agreement of the R^2 value, but this agreement is still too low for the proposed equation to fit the values well. In other words, the influence of the train speed can be neglected in this speed range and this data set.

Table 8 shows the influence of the days in operation between 11 and 22 November. The R^2 value is low and this means that the proposed model (Eq. 4) does not fit these values with good agreement, but the results indicate that the behaviour of the parameters and the different units differs. For the near unit the parameter Sd decreases, Sh increases, qR decreases and TH increases with time, which is in accordance with previous experience of the behaviour of the parameters [18]. The results for the far unit indicate the opposite for the parameters Sd and Sh, which is not in accordance with previous knowledge. Furthermore, the results for the far unit for TH indicate a large residual and the information can contain an outlier or an error.

Table 6 Wheel measurements from the far and near unit of the WPMS

ŷ	X	β_0	β_1	S	R^2 (%)	β_1/β_0 (%)
Sh	Near = 1	31.537	1.020	1.88504	6.9	3.2
Sd	Near = 1	28.325	−0.327	1.51548	1.17	1.2
qR	Near = 1	10.127	0.690	0.995364	9.7	6.8
TH	Near = 1	0.445	0.027	0.591508	0.1	6.1

Table 7 The influence of the train speed on the measurements of the far and near unit of the WPMS

ŷ	x	β_0	β_1	S	R^2 (%)
Sd N	Speed	31.51	−0.0556	1.34270	9.10
Sd F	Speed	33.74	−0.0862	13.39	13.39
Sh N	Speed	32.09	0.0045	1.87245	0.03
Sh F	Speed	31.62	0.0043	1.95914	0.02
qR N	Speed	10.99	−0.0028	0.999218	0.05
qR F	Speed	11.00	−0.1400	1.01722	0.95
TH N	Speed	−0.048	0.0082	0.601067	1.09
TH F	Speed	0.240	0.0033	0.594191	0.15

Table 8 The influence of the days in operation on the measurements of the far and near unit of the WPMS

$\hat{y}$	x	β_0	β_1	S	R^2 (%)
Sd N	Days	28.070	−0.0161	1.40736	0.13
Sd F	Days	27.518	0.1776	1.58986	10.65
Sh N	Days	32.289	0.0194	1.87174	0.11
Sh F	Days	31.620	−0.058	1.95118	0.83
qR N	Days	10.909	−0.203	0.997340	0.42
qR F	Days	10.339	−0.0466	1.01192	1.98
TH N	Days	0.384	0.0196	0.601141	1.06
TH F	Days	0.435	0.0023	0.594606	0.01

5.4 Discussion

The results from the case study are discussed below.

5.4.1 Level of Accuracy

With regard to the comparative accuracy of the two measurement units, the near unit has higher accuracy due to the smaller standard deviation for all the wheel profile measurements. Six different auto-generated measurements from the near unit show that the Sh measurements range from 33.2 to 33.5 mm, the Sd measurements from 25.6 to 26.6 mm, and the qR measurements from 10.1 to 10.4 mm. Consequently, the accuracy of the WPMS can be defined for Sd as being ±0.42 mm and for Sh as being ±0.08 mm according to the variation of wheel roundness [21]. This defined accuracy applies for optimal conditions and in this case for six samples for one wheel for the near unit.

5.4.2 Differences Between the Far and the Near Unit

The far unit has a lower process rate than the near unit. Regarding the far unit, seven out of 40 measurements were not usable due to missing data, whereas regarding the near unit, all the 40 measurements were useable. Furthermore, there is a significant difference between the results of the far and near units. The far unit has a larger standard deviation for all the wheel measurements (Sd, Sh and qR, see Fig. 1). The measurement parameter with the largest standard deviation is Sd, and the reason for this can be that this measurement is estimated from two different pictures.

With a CI of 95 %, there are differences between the far and the near measurement units of the WPMS for the wheel parameters Sh, qR, and TH; for the Sd, this hypothesis fails, meaning that there is a significant difference between the far and the near unit for this given CI.

5.4.3 Comparing the WPMS with the Manual Equipment

The comparison between the wheel parameter measurements from the WPMS and the manual measurements shows a difference concerning Sd of around 7 % and a difference concerning TH of around 5 %. The wheel profile shape obtained with the WPMS and that obtained with the manual measurements differ from each other.

5.4.4 Regression Analyses

The R^2 values from the regression analyses are rather weak and the data set seems to be too small to run linear regression analyses and obtain a strong R^2 value. Therefore, the influence of the passing train speeds between 50 and 72 km/h on the measured values is negligible; the same applies to the influence of wear due to these 10 days in operation. However, Table 8 shows that the results for the influence of the days in operation differ between the far and near unit with regard to Sd and Sh and indicate different behaviour.

5.4.5 Further Analyses Based on the Data

The data differ between the far and the near unit. Before calibration or some other error correlation, the data are not useful as a basis for further calculations and predictions.

6 Conclusions

The proposed method for data assessment can be used to explore and describe the basic quality features of condition monitoring data. The method can be adapted for condition monitoring systems in other fields of application than railway applications such as WPMS.

According to the results for the data assessed in the presented case study, these data are not useful or reliable for further engineering simulation, analysis or direct decision support without calibration or error correction of the far unit. There are so many benefits to be derived from the WPMS by both the infrastructure manager and the train operators, if the quality issue of the system is addressed. For instance, reliable alarms could trigger decisions and actions when some threshold has been passed. Further work needs to contain an investigation of a quality test algorithm which can increase the reliability of the data. Finally the regression analysis needs either a more developed model or a larger data set to fit the proposed model well.

Acknowledgments The authors wish to thank Trafikverket (the Swedish Transport Administration) for providing the data for this paper and Luleå Railway Research Center (JVTC) for financing this work. Last but not least, the authors wish to thank Jing Lin at LTU for useful and valuable advice for improving this article.

References

1. CEN (2010) EN 13306: maintenance terminology. European Committee for Standardization, Brussels
2. Ollier DB (2006) Intelligent infrastructure—the business challenge. The Institution of Engineering and Technology international conference on railway condition monitoring. The Centennial Centre, Birmingham, UK, pp 1–6
3. IEC (2010) IEC 60300-3-14: dependability management—part 3–14: application guide—maintenance and maintenance support. International Electrotechnical Commission (IEC), Geneva
4. American Bureau of Shipping (2004) Guidance notes on reliability-centred maintenance. American Bureau of Shipping, Houston
5. CEN (1999) EN 50126: railway specifications—the specification and demonstration of reliability, availability, maintainability and safety (RAMS). European Committee for Standardization, Brussels
6. Asplund M (2014) Wayside condition monitoring technologies for railway systems. Licentiate dissemination, Luleå University of Technology
7. Barke D, Chiu K (2005) Structural health monitoring in the railway industry: a review. Struct Health Monitor 4(1):81–94
8. Lagnebäck R (2007) Evaluation of wayside condition monitoring technologies for condition-based maintenance of railway vehicles. Licentiate Thesis, Luleå University of Technology, Luleå
9. Brickle B, Morgan R, Smith E, Brosseau J, Pinney C (2008) Identification of existing and new technologies for wheelset condition monitoring: RSSB report for task T607, TTCI (UK) Ltd., Rail Safety and Standards Board, London
10. Kahn BK, Strong DM, Wang RY (2002) Information quality benchmarks: product and service performance. Commun ACM 45(4):184–192
11. Yang Z, Cai S, Zhou Z, Zhou N (2005) Development and validation of an instrument to measure user perceived service quality of information presenting web portals. Inf Manag 42(4):575–589
12. Lee YW, Strong DM, Kahn BK, Wang RY (2002) AIMQ: a methodology for information quality assessment. Inf Manag 40(2):133–146
13. Pipino LL, Lee YW, Wang RY (2002) Data quality assessment. Commun ACM 45(4): 211–218
14. Berggren E (2010) Efficient track maintenance: methodology for combined analysis of condition data. Proc Inst Mech Eng Pt F: J Rail Rapid Transit 224(5):353–360
15. Utne IB (2012) A structured approach to improved condition monitoring. J Loss Prev Process Ind 25(3):478
16. Asplund M, Larsson D, Rantatalo M, Nissen A, Kumar U (2013) Inspection of railway turnouts using camera. World Congress of Railway Research, Sydney, pp 1–7
17. Trafikverket (2005) Safety inspections for permanent assets (Säkerhetsbesiktning av Fasta Anläggningar, BVF 807.2), Trafikverket, Borlänge
18. Braghin F, Bruni S, Lewis R (2009) Railway wheel wear. In: Lewis R, Olofsson U (eds) Wheel-rail interface handbook. Woodhead Publishing, Cambridge, UK, pp 172–219

19. Lin J, Asplund M (2013) Comparison study for locomotive wheels' reliability assessment using the Weibull frailty model. Eksploatacja i Niezawodnosc—Maintenance Reliab 16(2): 276–287
20. Iwnicki S, Björklund S, Enblom R (2009) Wheel-rail contact mechanics. In: Lewis R, Olofsson U (eds) Wheel-rail Interface Handbook. Woodhead Publishing, Cambridge, UK, pp 58–92
21. Fröhling R, Hettasch G (2010) Wheel-rail interface management: a rolling stock perspective. Proc Inst Mech Eng Pt F: J Rail Rapid Transit 224(5):491–497
22. Ekberg E, Kabo E, Kartttonen K, Lindquist B, Lundén R, Nordmark T, Olovsson J, Salmonsson O, Vernersson T (2013) Identifying root causes of heavy haul wheel damage phenomena. In: 10th IHHA conference, pp 520–528
23. Lin J, Pulido J, Asplund M (2014) Analysis for locomotive wheels' degradation. In: 2014 annual reliability and maintainability symposium (RAMS), pp 1–7
24. Deuce R (2007) Wheel tread damages—an elementary guide line, 100115000. In: Deuce R (ed) Bombardier Transportation GmbH, Netphen, Germany
25. Asplund M, Gustafsson P, Nordmark T, Rantatalo M, Palo M, Famurewa SM, Wandt K (2014) Reliability and measurement accuracy of a condition monitoring system in an extreme climate: a case study of automatic laser scanning of wheel profiles. Proc. Inst Mech Eng Pt F: J Rail Rapid Transit 228(6):695–704
26. Esveld C, Gronskov L (1996) MiniProf wheel and rail measurement. 2nd mini conference on contact mechanics and wear of rail/wheel systems, Budapest
27. Greenwood (2015) Webpage www.greenwood.dk/miniprof.php. Accessed: 17 Feb 2015