Decision Engineering

Louis Redding
Rajkumar Roy
Andy Shaw *Editors*

Advances in Through-life Engineering Services

Springer

Decision Engineering

Series editor

Rajkumar Roy, Cranfield, Bedfordshire, UK

More information about this series at http://www.springer.com/series/5112

Louis Redding · Rajkumar Roy
Andy Shaw
Editors

Advances in Through-life Engineering Services

Springer

Editors
Louis Redding
Cranfield University
Cranfield, Bedfordshire
UK

Andy Shaw
Cranfield University
Cranfield, Bedfordshire
UK

Rajkumar Roy
Department of Manufacturing & MSAS
Cranfield University
Cranfield, Bedfordshire
UK

ISSN 1619-5736 ISSN 2197-6589 (electronic)
Decision Engineering
ISBN 978-3-319-49937-6 ISBN 978-3-319-49938-3 (eBook)
DOI 10.1007/978-3-319-49938-3

Library of Congress Control Number: 2017934464

© Springer International Publishing AG 2017
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. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Some food for thought!

When you as a designer design something that burdens a community with maintenance and old world technology, basically failed developed world technology, then you will crush that community way beyond bad design; you'll destroy the economics of that community, and often the community socially is broken.

Cameron Sinclair

At its heart, engineering is about using science to find creative, practical solutions. It is a noble profession.

Queen Elizabeth II

A good decision is based on knowledge and not on numbers.

Plato

Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom.

Clifford Stoll

Any enterprise CEO really ought to be able to ask a question that involves connecting data across the organization, be able to run a company effectively, and especially to be able to respond to unexpected events. Most organizations are missing this ability to connect all the data together.

Tim Berners-Lee

Failure is central to engineering. Every single calculation that an engineer makes is a failure calculation. Successful engineering is all about understanding how things break or fail.

Henry Petroski

Design is a funny word. Some people think design means how it looks. But of course, if you dig deeper, it's really how it works.

Steve Jobs

When everything seems to be going against you, remember that the airplane takes off against the wind, not with it.

Henry Ford

A good scientist is a person with original ideas. A good engineer is a person who makes a design that works with as few original ideas as possible. There are no prima donnas in engineering.

Freeman Dyson

Aviation is the branch of engineering that is least forgiving of mistakes.

Freeman Dyson

Preface

Since the publication of the first book, "Through-life Engineering Services: Motivation, Theory and Practice" much has happened in the world relative to the service and support provision for complex engineering products. Continuing applied research in this area is being undertaken by a number of centres in the UK, Europe and the USA which seeks to develop innovative solutions and promote the exchange of ideas within a rapidly growing community of researchers, academics and industrial practitioners. This has seen progress being increasingly disseminated at the annual Through-life Engineering Services Conference (TESConf) which is hosted by Cranfield University in the UK. In alignment to the increasing international interest of the TES concept and scope TESConf 2017 will leave Cranfield and begin an international journey. As the conference starts it is international journey it will initially be hosted by the Bremen Institute for Mechanical Engineering at the University of Bremen in Germany.

The Through-life Engineering Services Centre at Cranfield University together with its co-collaborating research partners at Durham University (UK) have provided research focus and direction supported by funding from the Engineering and Physical Sciences Research Council and key industrial partners but is now moving to an industrial funding programme. This has demonstrated ongoing strong commitment from its key industrial supporters Rolls-Royce, BAE Systems, Babcock International, the UK Ministry of Defence and Bombardier Transportation. Through-life Engineering Services have also enjoyed support from the British Standards Institute who are now gathering momentum from industry for the development of a framework standard for through-life engineering services.

The most significant recent development has been the development and launch of a UK national strategy in Through-life Engineering Services. This is being led by an industrial and academic steering committee co-chaired by Rolls-Royce and the High Value Manufacturing Catapult. This initiative has captured the minds of a large number of key senior industrialists and is seen as a key milestone by academics and practitioners.

This book contains a compendium of contributions from leading international academics, researchers and practitioners who are continuing to develop Through-life

Engineering Services so as to provide aligned technical and business solutions for organisations seeking to compete through the adoption of ever-increasing service provision in support of their manufactured products. This publication builds upon the work presented in the first book and the accompanying body of literature. In Part I the chapters present the journey undertaken to realise the UK National Strategy for Through-life Engineering Services. They present the outputs from a series of workshops that were synthesised to create the final strategy document. Part II presents contributions relative to TES and the design function. In this section the relationship between warranty supporting the installed product base, and knowledge management are introduced. This is supported by the presentation of service support considerations undertaken when designing a civil aerospace gas turbine.

Part III goes on to discuss the role of data, diagnostics and prognostics within system design engineering for through-life engineering services and supporting complex systems which include both autonomy and design of contracts. The following section offers contributions from academia and industry dealing with how real systems and their components degrade. It looks at novel techniques for assessing such degradation and damage to help inform the replace or repair decision. Section V discusses further the importance of system design and presents a novel solution employing modelling techniques within the UK Rail Sector.

Building upon contributions within the previous book, Part VI presents contributions which address the important subjects of Cost Modelling, Planned Obsolescence and Contract Considerations. The final section of this work (Part VII) offers contributions which further address the importance of autonomous maintenance, self-healing and other emerging product support techniques.

There is no doubt that this field will continue to grow as we manufacture systems and products which are ever more complex and interconnected. The influence of this connectivity is already becoming relevant and as contracting mechanisms change and mature not only is the ownership of the product becoming less certain but also the ownership of the performance data it is producing in operation. We, the editors, look forward to many more years of interesting challenges in this new and exciting field.

Cranfield, UK

Louis Redding
Rajkumar Roy
Andy Shaw

Acknowledgements

The editors would like to acknowledge and thank for the guidance and support of all who contributed to the preparation and writing of this manuscript. The initial guidance relative to the scope of this text came from analysis of the outputs from a series of workshop events hosted by Cranfield University, which was facilitated by two of the co-editors of this work, Mr. A. Shaw and Prof. R. Roy. These events were attended by senior academics, industrial practitioners, researchers and consultants. This was supported by an analysis of trends within the literature at the time of preparation of the proposal for this work. To all those who participated the editors wish to express their gratitude.

Second, the editors wish to acknowledge and thank the 'unknown' peer reviewers who upon receipt of the book proposal gave their strong support of this project and gave constructive comments as to how the book should be structured and gave insight into the subject areas for inclusion.

The editors would like to express their sincere gratitude to all the contributing authors of the chapters herein who took time from their busy schedules to contribute to this work. Particular thanks are offered for the provision of their informed insight and support when constructing this compendium of chapters.

Finally, the editors wish to thank Springer Publishing for their proactive support throughout the undertaking of this work without which the contributions contained herein would not have been brought to the reader.

Contents

Part I Developing a Strategy for Through-life Engineering Services

Part II Through-life Engineering Services and Design

Part III The Role of Data, Diagnostics and Prognostics in Through-Life Engineering Services

Part IV Component Degradation and Design in Accelerated Life Testing Data

Part V System Degradation and Design

Part VI Cost, Obsolescence, Risk and TES Contract Design

Part VII Autonomous Maintenance and Product Support

Chapter 1
Introduction

Louis Redding

Abstract Through-life Engineering Services (TES) are continuing to develop. They are being adopted by manufacturing organisations at increasing rates as companies seek to move towards offering advanced service solutions in support of their product offerings. This chapter introduces the second book in a planned series of contributions by the editors. It presents further developments relative to the motivation, theory and practice relating to TES. The chapter offers to the reader a developing and strengthening rationale for the adoption of TES solutions. It presents the book structure and gives insight as to the methods used by the editors to identify the focus of study. Insight is given to the reader and those who will benefit from this compendium of contributions from eminent scholars and practitioners who are currently conducting state of the art research into TES, or applying its principles to achieve strategic advantage through the control and mitigation of risk.

1.1 Introduction to Through-Life Engineering Services

As manufacturing organisations seek to maintain and improve upon their strategic and competitive positions the adoption of Product Service Systems (PSS) [1–5] transitioned through the process of *servitization* [6–8] is becoming increasingly observed. The drivers for the adoption of PSS and *servitization* initiatives within the manufacturing organisation's business model comes from either (i) the need to combat increasing competition from low cost economies, (ii) the need to increase the sustainable offering in deference to ever increasing demand and finite resource, or (iii) a combination of the two. The origins of both concepts are discussed at great length within the literature as researchers and academics seek to further understand the contents and mode of application for each. It is also recognised that there is a

---

L. Redding (✉)
Cranfield University, Cranfield, Bedfordshire, UK
e-mail: louis.redding@cranfield.ac.uk

L. Redding
Jaguar Land Rover Plc, Gaydon, UK

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_1

Table 1.1 Main and sub categories of products and services [11]

Pure manufacturer	…..the organisation provides manufactured goods at a single point contract transaction with no additional service or manufacturer support. (The customer consumer adopts all the risks relating to product functionality and availability for use)
Product oriented PSS	…..the organisation provides manufactured goods and the ownership of those goods passes to the customer/consumer at the point of sale. Additional services are provided by way of service bundles ranging from base (parts and/or service) to intermediate (warranties)
Use oriented PSS	…..the manufacturer produces a product and retains ownership of that product and seeks to sell the use or function of the product to the consumer. This model appears in leasing contracts, but more complex long term, and through life availability contracts are emerging. With these contracts the manufacturer retains the risk relating to product function.
Results oriented PSS	….products are replaced entirely by services. This is seen in telephony (voice mail), IT (software products) and business services.
Pure service provider	….this is where there is no tangible product and the consumer purchases a pure service

servitization continuum (Table 1.1). Here one sees at either end of this continuum the organisation that is the 'pure manufacturer' or at the opposing end the 'pure service provider" [9]. Tukker et al. [10, 11] also identify several distinct stages along the scale which offer various increasing levels of service. Whilst the main focus within the literature is centred upon the manufacturing company wishing to add increasing levels of service to its customer offering it should not be overlooked that the pure service provider may wish to move towards the manufacturer by adding elements of production to its offering by acquisition or vertical integration. The point to remember is that transition along Tukker's continuum can go both ways.

Baines et al. [12, 13] in their research into the concept of PSS and *servitization,* and the strategy formulation process [14, 15] which enables a stakeholder to facilitate product aligned transformational change through the 'continuum', have further sought to clarify the levels of service that are offered at differing levels of customer support (Fig. 1.1). Here we start with the pure manufacturer who offers design and production of the product which is then sold as a single unsupported transaction. This entity is rare as consumer protection and optional warranty systems are, or are increasing becoming, the norm driven by customer and transactional protection legislation. As the organisation moves towards offering a service we see organisations operating at Baines's 'Base Level' of service offering no more than exchange equipment and spare parts as required.

The second stage, Baines calls the 'Intermediate Service' level. Here one sees the addition of maintenance, repair and overhaul (MRO) being offered in support of the product. If there are franchise service providers (i.e. car dealerships etc.) aligned to this post sale support then training of maintenance teams is also provided by the Original Equipment Manufacturer (OEM) to ensure quality of MRO support and

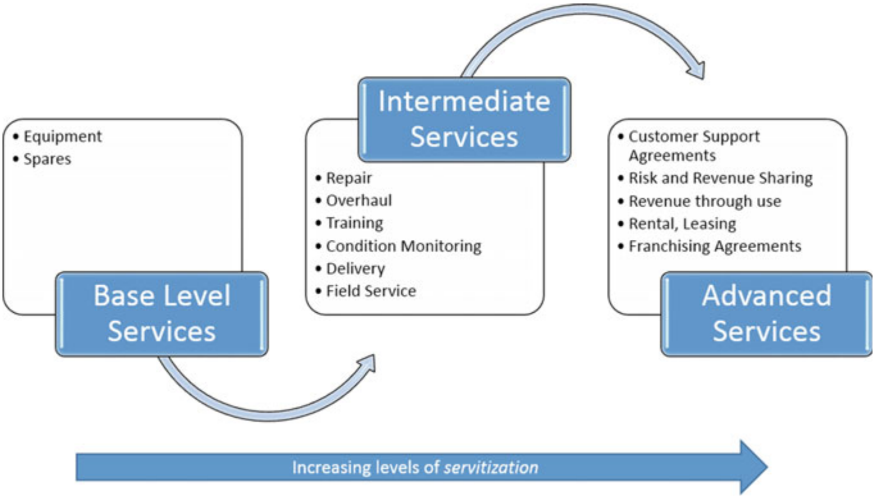


Fig. 1.1 Increasing levels of service as organisations '*servitize*'

accompanying field service either by the franchisee or third party service providers and product support organisations [e.g. Automobile Association (AA) and Royal Automobile Club (RAC)]. It is at this stage we see the emergence of condition based monitoring in support of the product during use. This is seen as the base line for the emergence of rudimentary TES solutions.

The final stage defined by Baines is that of "Advanced Services". At this level we see support agreements and risk to revenue agreements emerging. Typically, the products supported are high value engineering products (aeroplanes, trains, ships, cars, machine tools etc.). The developing business models which seek to acquire revenue from the use of the product through operating franchise agreements (for example Airlines and Train Operating Companies) and latterly availability contracting are facilitated by these advanced services.

Building upon Baines's research and that of the author into formulation methodologies which enable manufacturing organisations make the transition to ever increasing servitized product offerings the author introduced a strategy formulation methodology for companies seeking to compete through Integrated Vehicle Management Systems (IVHM) enabled service delivery systems [16]. Expanding upon the knowledge gained in compiling the previous edited book on this subject perhaps a more aligned label for Tukker's Continuum, and Baines's Advanced Services when acknowledging that these service offerings are provided throughout the product's life-cycle is:

Through-life Engineering Services enabled Service Delivery Systems.

In the first book the author introduced the reader to TES by way of a compendium of contributions from leaders in the field and through analysis of the literature at the time, he searched for a definition for TES which fulfilled the requirements for definitions. Namely, for a definition to be meaningful it should have *dimension, application, purpose* and *context*. After a thorough review of literature relative to TES and closely aligned concepts conducted between 2008 to the time of publication of the first book (2015), the following initial definition was offered:

> Through-life Engineering Services are…the…[*product of the*]… application of explicit and tacit 'in-service knowledge' supported by the use of monitoring, diagnostic, prognostic, technologies and decision support systems …[*which are applied*]…. whilst the product is in use, and…[*or during*]…..maintenance, repair, and overhaul (MRO) functions to mitigate degradation, restore 'as design' functionality, maximise product availability, thus reducing whole life-cycle cost. [17]

Since this definition was offered there has been little progress which sought to evolve this contribution whilst meeting the criteria of robustness. Namely that of meeting the elements of dimension, application, purpose, and context. The latest ones to be offered were during the launch of the TES National Strategy held at the Institution of Engineering and Technology, London (2016), and in the advertising literature for the British Standards Institution (BSI) and the TES Centre's invitation to participate in developing "a proposal for a new framework standard for sharing best practice in Engineering Services" which was held at Cranfield in 2016.

They being:

> Through-life Engineering Services (TES) is a 'whole life' product support strategy that ensures that the product's function is available for use throughout the product's design life. The concept covers all life stages from conception, through design, manufacture, operational use, to final end of life disposal. [Professor Raj Roy, London July: 2016]

And…….

> Through-life engineering services, (TES) comprise the design, creation and in-service sustainment of complex engineering products with a focus on their entire life cycle, using high-quality information to maximise their availability, predictability and reliability at the lowest possible through-life cost. [Roy and Sheridan: Press, July: 2016]

In presenting these three definitions the author notes that they are generally aligned and when ones reads them together it is felt that they collectively offer an identity for Through-life Engineering Services which meets the dimensions of a definition as previously discussed. There still remains however no universally accepted definition for TES but the author remains optimistic that this may appear once an international standard (ISO) appears.

1.2 Understanding the Target Audience for This Book

As discussed in the first book of this series, this compendium of work is suited to academics, researchers, post-graduate students, and industrial practitioners who are seeking to understand the principles and motivation behind the adoption of TES solutions. Typically, those studying, conducting research, or working in the following arenas will find this book of interest:

- Maintenance, Repair and Overhaul (MRO) activities which are either stand alone in support of existing product offerings, or part of an integrated advanced service offering within a Product Service System (PSS)
- Life-cycle Engineering (LCE)
- Life-Cycle Cost Engineering
- Asset Management (AM)
- Product Life-cycle Management (PLM)
- Design for Service
- Warranty and Quality Management and Engineering
- Risk Mitigation
- Original Equipment Manufacturers (OEM's) who are offering, or seek to offer Product-Service Solutions.
- Maintenance, Repair, and Overhaul (MRO) Organisations or individuals working through the supply chain in support of Product-Service Solutions.
- Those engaged in the development and execution of Availability Contraction
- Consultants operating in the field of Product Service Systems and the maintenance arena.

In addition to those listed above, this book could be of interest and a valuable resource to postgraduates who are studying for Postgraduate qualifications in any of the following areas:

- MSc Through-life Engineering Services
- MSc Engineering Management
- MSc Quality/Warranty Management
- MSc Maintenance Engineering and Asset Management
- MSc Maintenance Engineering
- MSc Maintenance Management
- MSc Aircraft Management Systems
- MSc Through-life Systems Sustainment
- MSc Integrated Vehicle Health Management
- MSc/PGCert Cost Engineering
- Any MSc Course relating to Maintenance, Repair and Overhaul (MRO)
- Any MSc Course relating to engineering and risk elements of Product-Service Systems.

1.3 Developing the Structure for This Book

In seeking to derive the structure for the first volume [18] of this proposed series the author sought to address the following questions:

i. What areas of TES should be discussed?
ii. How is the decision made as to what to include and exclude from any emerging themes?
iii. How to select the authors who contribute to the text?
iv. How to align the work to current and emerging research interests and the requirements of practitioners?
v. Who will be the potential readers of the text and how the contribution can meet their needs?
vi. What level of science should this work address? [18].

The process is fully described in the first chapter of the previous publication [18] with the methodology employed to ascertain the emergent themes being discussed. The themes for the publication were defined after an analysis of the data and findings from industrial surveys, a state of the art literature review, and practitioner workshops. The final themes that were identified from this analysis and used in the publication were:

- Data, Diagnostics and Prognostics,
- Component Degradation and Design,
- System Degradation and Design,
- Cost, Uncertainty, Risk, and Standards,
- Autonomous Maintenance.

Building upon the foundations laid within the first book the author adopted the same rhetorical questions as above to inform the structure of this work. In seeking to answer these questions the following data sources where referred to:

- Published Proceedings of the 1st International Conference on Through-life Engineering Services [19]
- Published Proceedings of the 2nd International Conference on Through-life Engineering Services [20]
- Published Proceedings of the 3rd International Conference on Through-life Engineering Services [21]
- Published Proceedings of the 4th International Conference on Through-life Engineering Services [22]
- EPSRC Centre for Innovative Manufacture Through-life Engineering Services —Annual Report (2014-2015) [23]
- Interim Report on TES National Strategy Workshops (2016) [24].

Firstly all of the contributions to the literature documented in the Conference Proceedings (2012–2015) were listed, analysed, and mapped into themes and detailed sub groups. Also keynote papers that were presented at each conference were read and grouped in accordance with their focus and emerging theme. This generated 25 sub-groups which were then re-analysed during assessment of abstracts to generate a list of 11 themes which contained a significant contribution of papers. A summary of the findings is illustrated in Table 1.2.

Table 1.2 Analysis of TES Themes is Previous Publications

	Proceedings 1st conference 2012	Proceedings 2nd conference 2013	Proceedings 3rd conference 2014	Proceedings 4th conference 2015	1st edited book 2015
No. of papers	52	79	48	49	25
Design and manufacture for TES	X	X	X	X	X
TES operations and management	X				X
Service informatics, CBM, IVHM, diagnostics and prognostics	X	X	X	X	X
Maintenance, repair and overhaul (MRO)	X	X	X	X	
Autonomy and self health technologies	X	X	X	X	X
System design for TES (including NFF)		X	X	X	X
Life-cycle engineering and obsolescence		X	X	X	X
Cost		X	X	X	X
Standards			X		X
Human influences		X		X	
TES modelling solutions	X	X	X	X	X
Strategy				X	
Logistics				X	
Non destructive testing (including acoustics emission testing)				X	X
Augmented reality				X	
Advanced ICT, IOT, and cyber security				X	X

Upon review it can be seen that the major foci in these literature sources which together contained 253 papers/chapters from eminent scholars and leading practitioners in the field deals with:

- Design and Manufacture for TES
- Service Informatics
- Autonomy, Self-Healing, and Repair
- TES Modelling Solutions.

These are seen as being of major interest in relation to TES. Significantly, yet unsurprisingly, the development and understanding of component and system degradation, Maintenance Repair and Overhaul (MRO), Life-cycle planning and Obsolescence, and cost all continue to attract a very strong focus.

Interestingly the drive for standards appears later as the conference series develops. The author suggests that this is of significance as at the time of the first Conference several large manufacturing organisations where already applying TES generic solutions to their operations in the form of Integrated Vehicle Health Management (IVHM) [25–27]. As is often the case, technologies evolve and are adopted. It is only after they prove to be of strategic advantage do standards and codes of practice emerge. At this point the author puts forward the hypothesis that in the commercial world first past the post results in winner takes all. A manufacturer that has already established working codes of practice in TES generic enabled business models (e.g. 'Power by the Hour'™ and 'Total Care'™) [28, 29] is in a strong position when seeking to define industry standards. Whilst BSI are continuing to work in this area in conjunction with EPSRC Centre for Innovative Manufacturing in Through-life Engineering Services (Cranfield) there are no national or international standards for TES. It is therefore significant that any emergent standards will probably reflect the strategic interests of those who are first past the post. Also of interest is that following an extensive call within the literature at associated conferences, and media there has been little if any contribution relative to standards or codes of practice relating to TES in answer to the call for contributions to this book. The author suggests that until the pull for international standards reaches the point of critical mass whereby that 'pull' becomes a market driver in of itself for adoption, then TES must be still in the early stages of evolution with a coherent identity still forming.

It is also important to note that interest in the links between ICT, Internet of Things (IOT), Cyber Security, and Augmented Reality are also emerging in the literature as presented at the TES Conference (2015). This was presented as a future development in the final chapter of the previous book and is seen by the author as the doorway to a paradigm shift in the design, manufacture, and support of '*informated*' (and autonomous) products in an ever increasing future web based connected society.

Finally, it is important to note that there appears to be an awakening within the literature directly related to TES aligned issues that appear with the human interface. Whilst contributions are only just appearing this interface will become of major significance if TES is to become a widely adopted methodology. The author predicts that the issues relating to TES and '*Virtual Teams*' will become of particular interest.

The next stage was to review the literature themes which emerged from a series of workshops which were conducted with invited industrial practitioners. This data was analysed together with the outputs from a recent meeting of the TES National Strategy Steering Committee and a TES White Paper that is currently in draft. The three one-day workshops were held during 2015/6 in the UK Cities of Bristol,

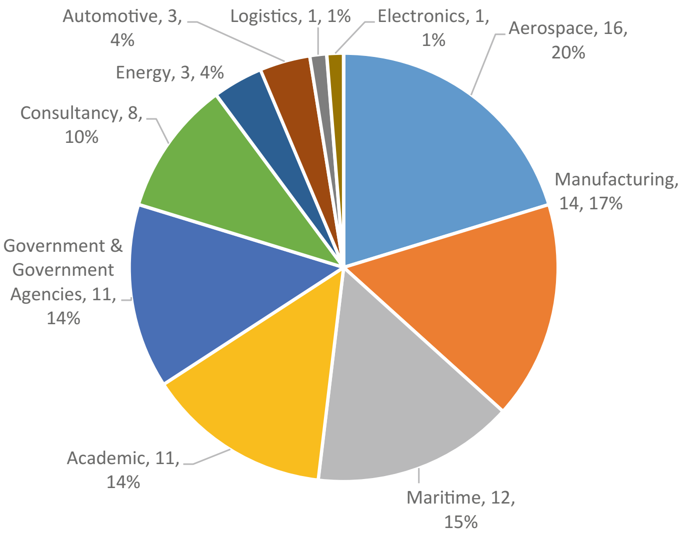


Fig. 1.2 Breakdown of practitioners attending TES workshops

Coventry and Glasgow and was attended by practitioners who came from several sectors (Fig. 1.2).

It is seen that the majority of those attending are from the Industrial Production Sector (Aerospace, Manufacturing, Maritime, Automotive, and Electrical) (61%) with Academics and Researchers and Government Agencies having a significant attendance (14%). Whilst this data is only an observation of those attending it is interesting to note that the distribution is relatively '*balanced*'. When one considers that TES and IVHM solutions appeared within the Aerospace sector it could be assumed that other 'sectors' appear to be awakening to the potential of TES applications.

It is also of interest to note the sectors that are absent from those listed above (e.g. Health, Insurance, Mineral Extraction, IT, Telcoms…etc.). In particular it is significant to note the non-attendance of the finance sector as this does not align to the interest and contributions as seen when reviewing the literature (Table 1.2). TES can be seen as a risk mitigation strategy for companies who generate revenue through availability contracting as successful application can maximise the products availability for use. It therefore follows that an understanding of cost models is required when applying TES or informing a future strategy for a TES enabled service delivery system. The absence of representatives from the finance sector is seen to be of significance when considering how the macro cost model might be defined.

Again the themes that emerged from the three workshops and an appraisal of the findings of the TES National Strategy Steering Committee and associated White Paper were tabulated and are presented in Table 1.3.

Table 1.3 Themes emerging from industrial workshops

	Bristol workshop	TES national strategy steering committee	TES white paper	Glasgow workshop	Coventry workshop
Culture and communication	**X**				
Collaborative mechanisms		**X**			
Technology and analytics			**X**		
Standards			**X**		
Supply chain issues			**X**		
Skills and training			**X**		
Export development	**X**				
Innovation				**X**	
Market knowledge					**X**
Design innovation					**X**
Institutional leadership					**X**
Productivity					**X**
Finance					**X**

It is seen that the views of the practitioners generally align with the findings within the literature. Whilst the vast majority of all contributions within the literature and presented at TES conferences thus far deal with the 'hard' aspects of engineering facilitating TES solutions it is seen from the above that 'soft' engineering and management issues are also emerging.

When seeking to bundle similar themes four top level themes emerge from the literature and practitioner engagement (Fig. 1.3). The hard engineering issues contain the literature and science relative to sensors, algorithms, system engineering, technology and analytics, design related tools and techniques, manufacture, communication science, condition monitoring, and an understanding of component degradation mechanisms and identification techniques. These elements of TES are well served within the body of knowledge and continue to evolve.

The 'soft' engineering focus includes aligned business models, cost modelling, simulation modelling etc., supported by the management focused disciplines of cost modelling, obsolescence management, business model formulation. Risk Management, Change Management, Culture and Communication, Virtual (Remote) Teams, and skills and development.

Fig. 1.3 Emergent dimensions of TES

As both the hard and soft dimensions develop there naturally follows a need for standards and governance the requirement for which is 'awakening' as manufacturing organisations start to engage with such bodies as the TES National Strategy Group and the British Standards Institute.

Finally, any solution needs to be fully aligned to the needs of the market and the stakeholders operating within the market. The need to understand this alignment is critical to the success of any TES initiative as demonstrated by the early adopters of TES and IVHM, open discussion across practitioners was emerging during the TES workshops held in 2016.

It is against the learning and knowledge gained from the aforementioned studies that the chapter submissions from academics, researchers and practitioners have been grouped into themes.

1.4 The Structure of This Book

Having reviewed all of the abstracts that were submitted in response to the chapter call and the following subsequent full chapters, the collective subject matter was grouped into themes in line with the learning from the process as described in Sect. 1.3. The result of this grouping defined the structure on this book with the following sections, namely:

- Part I: Developing a strategy for Through-life Engineering
- Part II: Through-Life Engineering and the Design Process
- Part III: The Role of Data, Diagnostics and Prognostics in Through-life Engineering Services
- Part IV: Component Degradation and Design in Through-life Engineering Services
- Part V: System Degradation and Design in Through-life Engineering Services

- Part VI: Cost, Obsolescence, Risk and TES Contract Design
- Part VII: Autonomous Maintenance and Product Support.

As in the first book in this series, each section as defined contains contributions from eminent scholars, researchers and industrial practitioners within the field of TES and its related technological applications. As was seen in the previous section, the number of active authors in the field is significant and increasing every year. The editors cannot include everyone or every focus and it is hoped that the contributions which are to be found in this second book of the series gives the reader further insights from which to continue to build upon their understanding of the principles and applications relative to TES. Finally, the editors hope that the contents herein will continue to promote both learning and development of the concepts.

References

1. Morelli N (2006) Developing new product service systems (PSS): methodologies and operational tools. J Clean Prod 14(17):1495–1501. doi:10.1016/j.jclepro.2006.01.023
2. Mont O (2000) Product-service systems. J Eng Des 20(4):327–328. doi:10.1080/09544820903149271
3. Baines TS, Lightfoot HW, Evans S, Neely A, Greenough R, Peppard J et al (2007) State-of-the-art in product-service systems. Proc Inst Mech Eng Part B J Eng Manuf 221 (10):1543–1552. doi:10.1243/09544054JEM858
4. Cavalieri S, Pezzotta G (2012) Product-service systems engineering: state of the art and research challenges. Comput Ind 278–288. doi:10.1016/j.compind.2012.02.006
5. Goedkoop MJ, Van Halen CJG, Te Riele HRM, Rommens PJM (1999) Product service systems, ecological and economic basics. Econ Aff 132. doi:10.1111/j.1365-294X.2004.02125.x
6. Baines T (2015) Servitization : the theory and impact. In: Proceedings of the spring servitization conference, p 312. ISBN 978 1 85449 492 4
7. Vandermerwe S, Rada J (1988) Servitization of business: adding value by adding services. Eur Manag J 6(4):314–324. doi:10.1016/0263-2373(88)90033-3
8. Baines T, Lightfoot H, Smart P (2011) Servitization within manufacturing. J Manuf Technol Manag 22(7):947–954. doi:10.1108/17410381111160988
9. Tukker A, Tischner U (2004) New business for old Europe. Suspronet, pp 132–148. http://www.ncbi.nlm.nih.gov/pubmed/18048240
10. Tukker A (2004) Eight types of product-service system: eight ways to sustainability? Experiences from suspronet. Bus Strat Environ 13(4):246–260. doi:10.1002/bse.414
11. Tukker A, Tischner U (2006) Product-services as a research field: past, present and future. Reflections from a decade of research. J Clean Prod 14(17):1552–1556. doi:10.1016/j.jclepro.2006.01.022
12. Baines TS, Braganza A, Kingston J, Lockett H, Martinez V, Michele P et al (2007) State-of-the-art in product service-systems. Proc Inst Mech Eng Part B J Eng Manuf 221(Part B):1543–1552. doi:10.1243/09544054JEM858
13. Ang G, Baines T, Lightfoot H (2010) A methodology for adopting product service systems as a competitive strategy for manufacturer. In: CIRP IPS2 Conference, pp 489–496. http://www.ep.liu.se/ecp/077/063/ecp10077063.pdf

14. Baines TP (2011) Service-led competitive strategies: an inaugural lecture by prof tim baines. Aston Business School. http://www.aston.ac.uk/aston-business-school/staff/academic/oim/tim-baines/
15. Redding LE (2014) A strategy for service delivery systems. Strategic Change, 23:287–302. doi:10.1002/jsc
16. Redding LE (2012) A strategy formulation methodology for companies seeking to compete through IVHM enabled service delivery systems. Cranfield University, Cranfield, Bedfordshire, UK
17. Redding LE (2015) Chapter 2: Through-life engineering services: definition and scope—a perspective from the literature. In: Redding LE, Roy R (ed) Through-life engineering services: motivation, theory, and practice, 1st edn. Springer; pp 13–28
18. Redding LE, Roy R (ed) (2015) Through-life engineering services: motivation, theory, and practice, 1st edn. Springer
19. Roy R, Shehab E, Hockley C, Khan S (ed) (2012). In: Proceedings of the 1st international conference on through-life engineering services. Enduring and cost-effective engineering support solutions. Cranfield University Press, Cranfield
20. Roy R, Tiwari A, Shaw A, Bell C, Phillips P (ed) (2013). In: Proceedings of the 2nd international through-life engineering services conference. Elsevier, Cranfield
21. Roy R, Tomiyama T, Tiwari A, Tracht K, Shahab E, Shaw A (ed) (2014). In: Proceedings of the 3rd international conference in through-life engineering services. Elsevier, Cranfield
22. Roy R, Tiwari A, Tracht K, Shehab E, Menhen J, Erkoyuncu JA, Tapoglou N, Tomiyama T (ed) (2015). In: Proceedings of the 4th international conference on through-life engineering services, Procedia CIRP. Elsevier, pp 1–288
23. Unknown (2015) Through-life engineering services: annual report 2014/5. Cranfield University Press, Cranfield, Bedfordshire, UK
24. Shaw A (2016) Interim report on TES national strategy workshops, unpublished draft. Cranfield University, Cranfield, Bedfordshire
25. Benedettini O, Baines TS, Lightfoot HW, Greenough RM (2009) State-of-the-art in integrated vehicle health management. Proc Inst Mech Eng Part G J Aerosp Eng 223(Part G):157–170. doi:10.1243/09544100JAERO446
26. Scandura P (2005) Integrated vehicle health management as a system engineering discipline. In: 24th Digital Avionics Systems Conference, vol 2, pp 7.D.1–1–7.D.1–10. doi:10.1109/DASC.2005.1563450
27. Esperon-Miguez M, John P, Jennions IK (2013) A review of integrated vehicle health management tools for legacy platforms: challenges and opportunities. Prog Aerosp Sci 19–34. doi:10.1016/j.paerosci.2012.04.003
28. Unknown (2012) Rolls-royce celebrates 50th anniversary of power-by-the-hour. Rolls Royce Media. http://www.rolls-royce.com/media/press-releases/yr-2012/121030-the-hour.aspx
29. Unknown (2013) 'Power by the hour': can paying only for performance redefine how products are sold and serviced? Wharton Unviersity of Pennsylvania. http://knowledge.wharton.upenn.edu/article/power-by-the-hour-can-paying-only-for-performance-redefine-how-products-are-sold-and-serviced/. Accessed: 9 Aug 2016

Part I

Developing a Strategy for Through-life Engineering Services

Chapter 2
The Development of a UK National Strategy for Through-Life Engineering Services: Rationale and Process

Andy Shaw and Paul Tasker

Abstract This chapter describes the journey of the development of the UK National Strategy for Through-life Engineering Services (TES) from initial discussions with industry as to the need, through the issuing of a number of sector strategies in the UK, which omitted any reference to TES, to the design of the strategy development workshops. Initial industrial soundings are detailed which supported the need for the UK National Strategy as is the creation of an emergent industrial momentum. The importance of coherence of vision is underlined by extensive work with the voluntary industrial steering committee. The methodology adopted to generate the launch event white paper and cross-sectoral industrial economic study is given in detail. Pilots, tests and the consequent evolution of the workshops are described together with the range and depth of industrial engagement with the process. The chapter begins with a description of the industrial engagement strategy of the EPSRC Centre for Innovative Manufacturing in Through-life Engineering Services from which the UK National Strategy was born.

2.1 Introduction

The EPSRC Centre for Innovative Manufacturing in Through-life Engineering Services was launched in July 2011 as a research collaboration between Cranfield and Durham Universities. Early on in its initial five year funded life it established extensive contacts with relevant industrialists through the establishment of a Think Tank to better guide research directions and outputs in order to achieve maximum impact. Over time it was observed that the UK government was developing a number of sector strategies, most notably in aerospace [1] which made little or no reference to the service and support of complex engineered products.

A. Shaw (✉) · P. Tasker
Cranfield University, Cranfield, England
e-mail: andy.shaw@cranfieldmanufacturing.com

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_2

The concept of generating enhanced value by combining the provision of complex engineered products with associated services has been well documented under a number of different headings including product service systems and servitisation [2–6]. More complex contracting arrangements including leasing of assets such as trains and aircraft have prompted the investigation of different business models, and these have been championed by some key customers such as the UK Ministry of Defence (MOD) with availability contacting for both Tornado and Typhoon aircraft [7].

With practice varying between sectors and best practice sometimes hidden in sector silos, there is an opportunity to bring coherent and joined up thinking to a wide range of UK industry where they may be able to gain significant commercial advantage. Early consultation with the industrial network created by Centre established the need for the development of a coherent strategy for the UK. The Centre, its advisory board and executive committee decided to pursue the development of a National Strategy in TES. This fulfills in part the Centre's remit to act as a National resource.

2.2 The Initial Journey

The industrial network of the Centre was brought together in mid-2014 at the Institution of Mechanical Engineers in London for a dinner where a series of sector leaders spoke of the growing importance of TES to their industries. Chris White MP co-chair of the all-party parliamentary manufacturing group, gave the keynote speech. Interest and momentum was generated in building closer ties between sectors and learning beat practice. This meeting identified the need for a coherent UK National Strategy in TES.

Following the dinner in London an event was organised in parallel to the 2014 TES conference in November. Seven key themes were identified at the dinner. These were Knowledge Management, Lifecycle Management and Lifecycle Extension, Technology Innovation, Supply Network, Human Resources, Socio-Technical issues and the Defence Context. The first four of these themes were selected for detailed discussion at the conference session. The findings of the session in the four key areas explored led to a growing belief among the industrial and academic participants that there was a need for a TES National Strategy.

In May 2015 a further and much larger workshop was held at the Royal Academy of Engineering in London to address this need and move the debate on. A group of about 70 senior engineers, business leaders and government representatives gathered to collate their views on TES, the market position, the market needs, the benefits and challenges, and the possible Future State that a national strategy could deliver.

The key questions addressed were:

- How important is TES capability to the UK?
- What are the opportunities for innovation and development in TES?
- What are the barriers for realising these opportunities, at pace?
- What steps might mitigate these barriers?

The EPSRC Centre commissioned Raj Mehta, a consultant and former Operations Director at Bombardier Transportation and General Manager at British Airways, to prepare a report on the international market for TES [8]. He commented that by 2025, the global market for maintenance, repair and overhaul (MRO) services in civil air will be $89 billion, whilst in the UK the repair and maintenance (RAM) industry as a whole, as summed over all the SIC codes relating to RAM, will be worth in excess of £35 billion.

He also pointed out that there is strong evidence that TES accounts for high-value employment with wage rates about one and a half times the average earnings in engineering. Raj explained he drew on resources from the Office for National Statistics, and defence and aerospace company annual reports to compile his data.

This report formed the backdrop to the workshop activities. Delegates were asked to share and summarise their most significant issues and experiences with TES, and then to write these points on colour-coded Post-It notes to populate a central "strategic framework" addressing the state of TES in 2030, the state in 2015 and the opportunities and actions to move to the future stare. The strategic framework was in the form of a template divided into four layers to assist the groups' thinking about the issues involved:

Market Context—What are the driving needs for change?
Value Context—Trends in regulation, standards or other policies
Value Capture—How is value captured from TES?
Value Creation—What are the key activities and capabilities?

This provided a comprehensive collection of comments from industry on where TES is, the actions required and barriers to progress, and a vision of how the TES industry in Britain could be. Delegates were invited to discuss each other's outputs and summary discussions in plenary were held. The main thematic points identified were:

- In future, customers will only buy services—product-only providers won't exist—leading to polarized manufacturing—between throw-away and circular economy. TES is vital for (the missing link for) sustainability and the circular economy
- The UK leads today. TES is mission critical for long term growth/export in high value manufacturing—If we don't continue to lead we'll lose
- Government contracts, especially infrastructure, energy and transport are critical and could provide a game changer for TES and competitiveness in the future. Government can lead with value for money in government contracting

- Need better cost models and data: need to move from "open-loop" to "closed-loop" business with multi-functional management and collaborative behaviors. Needs new skills, training
- TES creates multi-functional, high-value jobs (average wages in engineering services are one and a half times those in mainstream manufacturing (Raj Mehta)
- The importance of standards and regulation—need to be enablers and support cross-industry knowledge transfer (e.g. aero/defence, auto into nuclear/energy and transport, e.g. rail)
- Is it a race to the bottom?—we really need to define cost and value for TES.

The significant consensus was that a national strategy could help with all these main points: there was a big market that the UK would otherwise lose out on, and much to commend a national approach to maintaining the UK's own long-life, complex assets. A report on the event was produced and circulated to participants. A white paper was prepared based on the outputs of the workshop and this formed the basis for the next stage in the strategy development [9].

2.3 The Launch and Development of the Strategy

Over the summer of 2015 Rolls-Royce and the High Value Manufacturing (HVM) Catapult agreed to co-chair a national working group to bring about the development of this strategy. A launch event incorporating the publication of the white paper [9] and a call for participation was held at the Palace of Westminster in early September 2015 with key representation from industry, Members of Parliament, civil servants and academia. Sixty participants heard of the importance of TES from Sir Peter Gregson, Vice-Chancellor and Chief Executive of Cranfield University, Chris White MP and Barry Sheerman MP, co-chairs of the all-party parliamentary manufacturing group and a panel of industrial and research specialist. The panel comprised David Benbow of Rolls-Royce, Vaughan Meir and Alan Murdoch of BAE Systems, Rob Cowling of Bombardier Transportation, Mark Claydon-Smith of the Engineering and Physical Sciences Research Council (EPSRC), Dick Elsy of the HVM Catapult and Rajkumar Roy of Cranfield University.

At the launch a number of organisations agreed to contribute to a steering committee for the National Strategy development under the joint chairs of Rolls-Royce and the HVM Catapult. Organistions volunteering included the UK Ministry of Defence (MOD), UK Government Department for Business Innovation and Skills (BIS), Innovate UK, the EPSRC, BAE Systems, Babcock International, Bombardier Transportation, Siemens UK, Si2 Partners, the Manufacturing Technologies Association (MTA), Aerospace Defence and Security (ADS), and the Universities of Aston, Cambridge and Cranfield.

This steering committee convened for the first time in late October 2015 and set a challenging timetable for the strategy development with a target publication date of early summer 2016. It would go on to meet a further six times endorsing and directing the strategy development process. Three regional workshops were scheduled for the strategy development, the first being held in Bristol in early December 2015 with two further workshops in Glasgow and Ansty near Coventry taking place in January 2016.

The design of these workshops was evolutionary and as new key themes were identified these were added in so that delegates at the final workshop were working on as complete a picture of the issues as it was possible to generate. In total eighty delegates attended the workshops and their output formed the starting point for a detailed piece of analytic work which synthesized these outputs into the National Strategy. Four initial themes came from the white paper [9] and to these a further nine were identified across the three workshops. Table 2.1 lists the themes and their sources.

The steering committee recognised that the early work on the size of the potential TES market was rather limited and commissioned Professor Alan Hughes of the Judge Business School at Cambridge University to undertake a more extensive study into the national economics of TES in parallel to the workshops. This study also included a survey of, and structured interviews with knowledgeable sector experts to validate the reach and influence of potential TES markets. Two reports were prepared, one concentrating on the economic metrics [10] and the other reporting on the sector interviews [11].

Table 2.1 Themes for the strategy and their source

Theme	Source
Culture and communications (including 'why TES'?)	Bristol workshop
Collaboration mechanisms	TES national strategy steering committee
Technology and analytics	White paper
Standards	White paper
Supply chain	White paper
Theme	Source
Skills	White paper
Export development	Bristol workshop
Innovation	Glasgow and coventry workshops
Market knowledge	Coventry workshop
Design innovation	Coventry workshop
Institutional leadership	Coventry workshop
Productivity	Coventry workshop
Finance	Coventry workshop

2.4 The Presentation of the Strategy and the Next Steps

The national strategy [12], along with the extensive report on TES markets [10] were presented at an event held at the IET in London and introduced by Chris White MP in early July 2016. Around one hundred delegates heard presentations on the strategy, its development, the potential markets and the future of research in this field. The development team formed a panel and answered questions on the reports which were distributed to the delegates at the end of the event.

The team called for volunteers to help form a sector council in TES and a number of individuals from attending organisations have stepped forward. The UK MOD have ordered a print run of 100 copies of the strategy document to issue to their staff and supply and support chains. Work is now ongoing to stand up the council which will have its first meeting in the Autumn of 2016.

Acknowledgements The authors wish to acknowledge the support of the EPSRC Centre for Innovative Manufacturing in Through-life Engineering Services in supporting the development of this strategy.

References

1. Aerospace Growth Partnership (2012) Reach for the skies: a strategic vision for uk aerospace. Department for business, innovation and skills. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/31807/12-954-reach-skies-strategic-vision-uk-aerospace.pdf. Accessed 23 Aug 2016
2. Johnstone S, Dainty A, Wilkinson A (2009) Integrating products and services through life: an aerospace experience. IJOPM 29(5):520–538
3. Smith L, Maul R, Ng I (2014) Sevitization and operations management: a service dominant-logic approach. IJOPM 34(2):242–269
4. Baines T, Lightfoot HW (2013) Servitization of the manufacturing firm: exploring the operations practices and technologies that deliver advanced services. IJOPM 34(1):2–35
5. Lightfoot HW, Baines T, Smart P (2013) The servitisation of manufacturing: a systematic literature review of independent trends. IJOPM 33(11/12):1408–1434
6. Masood T, Roy R, Harrison A et al (2015) Integrating through-life engineering service knowledge with product design and manufacture. Int J Comput Integr Manuf 28(1):59–74
7. Hockley CJ, Smith JC, Lacey L (2011) Contracting for availability and capability in the defence environment, In: Ch13 in Ng I, Parry G, Wild PJ et al (eds) Complex engineering service systems: concepts and research. decision engineering series. Springer, London
8. Mehta R (2015) UK support and service industry: a high value employer and a net exporter. EPSRC centre for innovative manufacturing in through-life engineering services. http://www.through-life-engineering-services.org/downloads/Final_UK_Support_and_Service_Industry.pdf. Accessed 23 Aug 2016
9. Tasker P (2015) Making things work engineering for life—developing a strategic vision. EPSRC centre for innovative manufacturing in through-life engineering services. http://www.through-life-engineering-services.org/downloads/27302_White_paper_amended.pdf. Accessed 23 Aug 2016

10. Hughes A, Hughes J (2016) Through-life engineering services (TES): market and data review. EPSRC centre for innovative manufacturing in through-life engineering services. http://www.through-life-engineering-services.org/downloads/Market_and_Data_Review.pdf. Accessed 23 Aug 2016
11. EPSRC Centre for Innovative Manufacturing in Through-life Engineering Services (2016) Making things work better for longer: sector report. http://www.through-life-engineering-services.org/downloads/TES_Sector_Report.pdf. Accessed 23 Aug 2016
12. EPSRC Centre for Innovative Manufacturing in Through-life Engineering Services (2016) A national strategy for engineering services: delivering UK economic growth by making things work better for longer. http://www.through-life-engineering-services.org/downloads/Engineering_Services_Strategy_new.pdf. Accessed 23 Aug 2016

Chapter 3
The Development of a UK National Strategy for Through-Life Engineering Services: Workshop Outputs Analysis and Final Strategy Creation

Andy Shaw and Paul Tasker

Abstract This chapter gives in detail the data collected from the industrial workshops. It also covers the parallel work strand to validate and develop an economic case for sector focus on Through-life Engineering Services. Analysis methodologies are described and the resulting collated outputs listed. Next the drafting process is developed and stages of consultation and revision given. The search for a publishing authority within the UK government is discussed as are the invitations to significant figures for endorsement. Finally the strategy document is described and its key recommendations listed. The full strategy is listed as an appendix to this chapter for ease of reference.

3.1 Introduction

The previous chapter outlined the journey that has taken place to develop the National Strategy for Through-life Engineering Services. This chapter will cover the objectives of each workshop, their design and operation and the output results. Over 140 separate organisations took part in the National Strategy development with many sending multiple delegates to the various workshops and events.

A. Shaw (✉) · P. Tasker
Cranfield University, Cranfield, England
e-mail: andy.shaw@cranfieldmanufacturing.com

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_3

3.2 The Initial Workshop at the Royal Academy of Engineering

3.2.1 *Objectives, Activities and Templates*

The objective of this workshop was to gather industry, academia and government's views on the developing area of Through-life Engineering Services (TES). Delegates were asked to identify key issues and themes for development over the next fifteen years. A secondary but extremely important objective was to gain support for the concept of a National Strategy. The workshop drew on a wide range of companies, trade associations, and academics and a select group of key government departments.

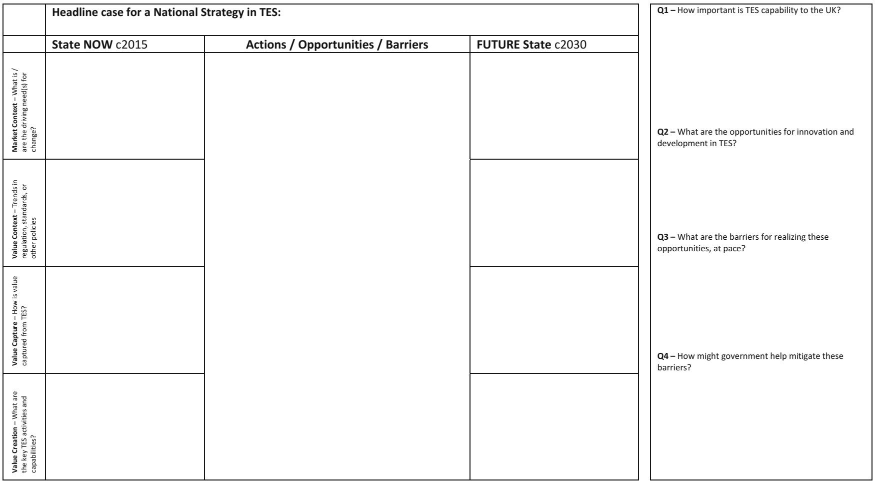
The template the teams were asked to populate is shown in Fig. 3.1.

3.2.2 *Direct Workshop Outputs*

Several delegates made general observations which commented on how mature the TES discipline has become and they were impressed with the high level of senior engagement.

A number remarked that TES:

- Should be a de facto method for managing scarce resources and extending product life. Reference was made to TES being an essential or key enabler for the circular or sustainable economy of the future.



	Headline case for a National Strategy in TES:		
	State NOW c2015	**Actions / Opportunities / Barriers**	**FUTURE State** c2030
Market Context – What is / are the driving need(s) for change?			
Value Context – Trends in regulation, standards, or other policies			
Value Capture – How is value captured from TES?			
Value Creation – What are the key TES activities and capabilities?			

Q1 – How important is TES capability to the UK?

Q2 – What are the opportunities for innovation and development in TES?

Q3 – What are the barriers for realizing these opportunities, at pace?

Q4 – How might government help mitigate these barriers?

Fig. 3.1 Template used at the royal academy of engineering workshop

- Needs multi-skilled people, engineering and other skills, and could be a means to up-skill the UK industrial workforce.
- While the UK appears to lead in some areas, more needs to be done to develop this leadership position in order to grow the UK share of the global market.
- There is an opportunity to accelerate capability growth and competitiveness by leveraging government and public sector "TES activities" particularly for the maintenance, refurbishment and improvement of UK infrastructure and this could be facilitated by government leadership—government demonstrating they were a service customer.

Although a few questioned how a national strategy might help, the significant consensus was that such an approach could help with all these main points: there was a big market that the UK would otherwise lose out on, and much to commend a national approach to maintaining the UK's own long-life, complex assets.

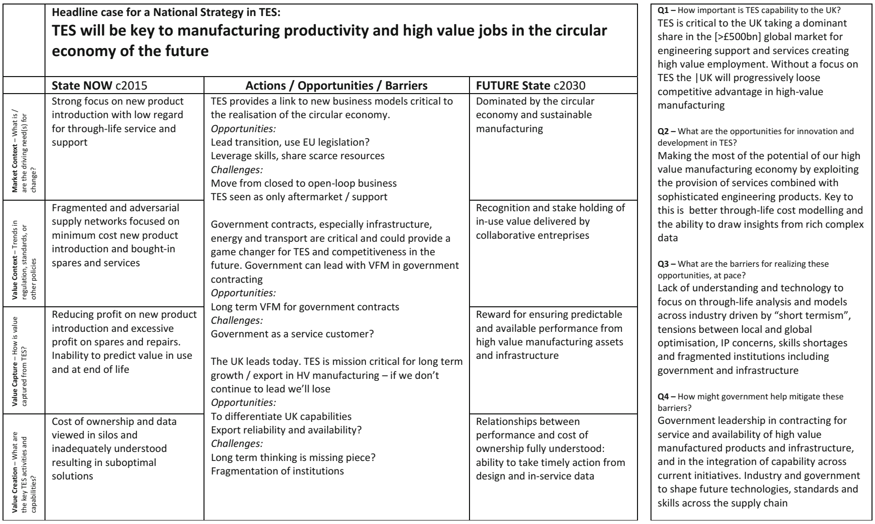
The final plenary discussion was based on the main key thematic points that had been identified during discussion of the poster presentations:

- In future, customers will only buy services—product-only providers won't exist —leading to polarized manufacturing—between throw-away and circular economy. TES is vital for (the missing link for) sustainability and the circular economy
- The UK leads today. TES is mission critical for long term growth/export in high value manufacturing—If we don't continue to lead we'll lose
- Government contracts, especially infrastructure, energy and transport are critical and could provide a game changer for TES and competitiveness in the future. Government can lead with value for money in government contracting
- Need better cost models and data: need to move from "open-loop" to "closed-loop" business with multi-functional management and collaborative behaviors. Needs new skills, training
- TES creates multi-functional, high-value jobs (average wages in engineering services are one and a half times those in mainstream manufacturing (Raj Mehta))
- The importance of standards and regulation—need to be enablers and support cross-industry knowledge transfer (e.g. aero/defence, auto into nuclear/energy and transport, e.g. rail)
- Is it a race to the bottom?—we really need to define cost and value for TES.

These main points are combined into a Summary Landscape in Fig. 3.2. This is the product of the comments from the strategy posters developed independently by all nine groups distilled into a single summary table. It reflects the most common and most important points and conclusions from the entire workshop.

Discussing these points in plenary, delegates commented:

Commoditisation of TES—"If the future vision is that we will build mostly things that are very serviceable, by creators, suppliers and competitors, companies may go through this cycle: a *trough of despair* and frantic competition with people trying to reverse engineer each other's products—then a *slope of enlightenment to*



	Headline case for a National Strategy in TES: **TES will be key to manufacturing productivity and high value jobs in the circular economy of the future**		
	State NOW c2015	**Actions / Opportunities / Barriers**	**FUTURE State** c2030
Market Context – What is / are the driving need(s) for change?	Strong focus on new product introduction with low regard for through-life service and support	TES provides a link to new business models critical to the realisation of the circular economy. *Opportunities:* Lead transition, use EU legislation? Leverage skills, share scarce resources *Challenges:* Move from closed to open-loop business TES seen as only aftermarket / support Government contracts, especially infrastructure, energy and transport are critical and could provide a game changer for TES and competitiveness in the future. Government can lead with VFM in government contracting *Opportunities:* Long term VFM for government contracts *Challenges:* Government as a service customer? The UK leads today. TES is mission critical for long term growth / export in HV manufacturing – if we don't continue to lead we'll lose *Opportunities:* To differentiate UK capabilities Export reliability and availability? *Challenges:* Long term thinking is missing piece? Fragmentation of institutions	Dominated by the circular economy and sustainable manufacturing
Value Context – Trends in regulation, standards, or other policies	Fragmented and adversarial supply networks focused on minimum cost new product introduction and bought-in spares and services		Recognition and stake holding of in-use value delivered by collaborative entreprises
Value Capture – How is value captured from TES?	Reducing profit on new product introduction and excessive profit on spares and repairs. Inability to predict value in use and at end of life		Reward for ensuring predictable and available performance from high value manufacturing assets and infrastructure
Value Creation – What are the key TES activities and capabilities?	Cost of ownership and data viewed in silos and inadequately understood resulting in suboptimal solutions		Relationships between performance and cost of ownership fully understood: ability to take timely action from design and in-service data

Q1 – How important is TES capability to the UK?
TES is critical to the UK taking a dominant share in the [>£500bn] global market for engineering support and services creating high value employment. Without a focus on TES the |UK will progressively loose competitive advantage in high-value manufacturing

Q2 – What are the opportunities for innovation and development in TES?
Making the most of the potential of our high value manufacturing economy by exploiting the provision of services combined with sophisticated engineering products. Key to this is better through-life cost modelling and the ability to draw insights from rich complex data

Q3 – What are the barriers for realizing these opportunities, at pace?
Lack of understanding and technology to focus on through-life analysis and models across industry driven by "short termism", tensions between local and global optimisation, IP concerns, skills shortages and fragmented institutions including government and infrastructure

Q4 – How might government help mitigate these barriers?
Government leadership in contracting for service and availability of high value manufactured products and infrastructure, and in the integration of capability across current initiatives. Industry and government to shape future technologies, standards and skills across the supply chain

Fig. 3.2 Consolidated output from all teams at the royal academy of engineering workshop

maturity in which some people will be good at maintaining, others design and others repair and companies find their niche [place in the value chain]."

Where is the cash?

Question: "TES is about a high value product which generates more jobs and yet costs less. So are we being so much more efficient on both sides of the contract without actually generating a lot of extra money? I can see senior accountants struggling to see the payback."

Response: If you do TES properly you can optimise both ends; minimise costs (especially material) and maximise revenue. Where else can you do this?

Maturity of the conversation, need to make Services a more attractive concept—The whole concept is moving on and accepted as a main theory. Now the challenge is 'how do you make this sexy'? Because often service and support is something you didn't do because often you were more interested in designing the product in the first place.

TES and the circular economy—Consider that future model where half the world is throwaway and half are assets that last indefinitely—what does that mean for manufacturing? The connection of the circular economy and sustainability with TES.

TES is not a panacea—In defence there is a lot of transportation, so it makes managerial sense to compartmentalise the purchase with the service. A lot of these industries are transport-based, freight based industries, but an example like hospital PPIs were a disaster for service and product buyers.

Understand where TES adds value—For TES to be successful, it's not about boasting about the UK's capability but more where the capability will be used and add value.

Base level education—Engineers that we employ have probably never heard of the concept of designing anything for maintenance. If you think services are 30–40% of total revenue opportunity in engineering, then where is the 30–40% in a typical engineering training course?

Based on the outputs from this workshop a white paper was written [1] and presented at the National Strategy kick-off event at the Houses of Parliament.

3.3 The Three Regional Strategy Development Workshops

3.3.1 Objectives, Activities and Templates

The objectives of these workshops were to broaden the engagement in the National Strategy development by engaging more companies and organisations from around the UK and getting them to contribute to the process in some detail. In this way they would be able to make a significant contribution to the development of the Strategy.

Four key themes had been spotted in the first workshop and by the initial discussions of the Steering Committee. The process was so designed that new themes could be identified and added into the process for further analysis by subsequent workshops.

Workshop attendees were first asked to populate a descriptive template of the TES landscape with noun/verb pairs, write them on a Post-it® and add them to it in the relevant phase or area of the diagram. This process was designed to get the delegates thinking about the fundamentals of the TES activity. Figure 3.3 shows the

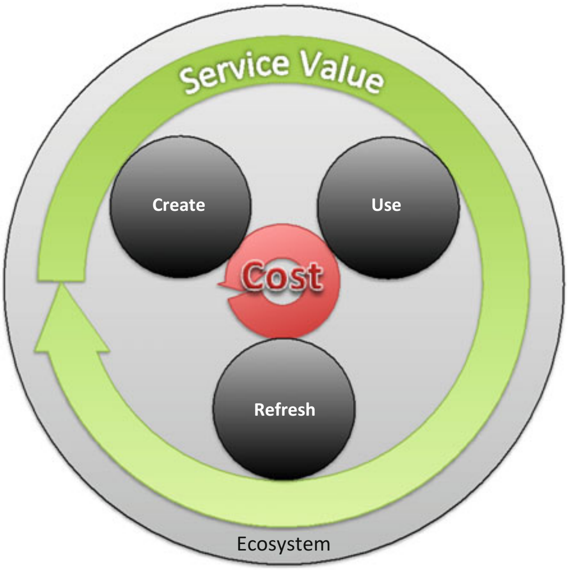


Fig. 3.3 The descriptive template of the TES landscape

model. The template contained six boxes each covering one aspect of the model: service value, cost, create, use, refresh and ecosystem.

After this brainstorming session the teams were asked to cluster the results and create summary topics within each sheet—giving each theme a title. Each delegate then voted using different coloured dots (max three red and three green dots per delegate) to identify priority areas. Green dots denoted areas of strength for the UK, red dots areas of weakness.

Delegates were then asked to identify the 'top five' and one 'wildcard' priority area and write each on a red or green (as appropriate) Post-it®. These were then added to the appropriate theme where possible. Where no relevant theme existed delegates were invited to suggest a new theme.

Teams were then allocated themes to develop on a new template. Where there were more teams than themes teams were allocated duplicate themes and asked to compare notes after separate analysis to identify key issues at the plenary feedback stage. The theme template is shown in Fig. 3.4. Teams were asked to work through the boxes in numerical order starting with "1 Vision", then "2 Current state" and so on.

So under the Vision heading teams were asked to describe the future (2030) world, for example: market place conditions, common products and service offerings, the state of Science and Research, and general competitiveness of UK Industry with particular regard to the theme being considered.

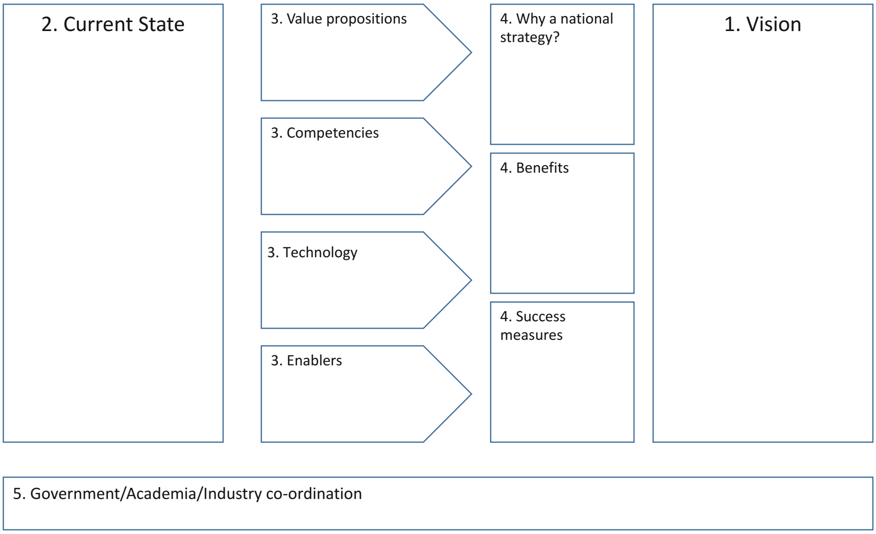


Fig. 3.4 Theme development template

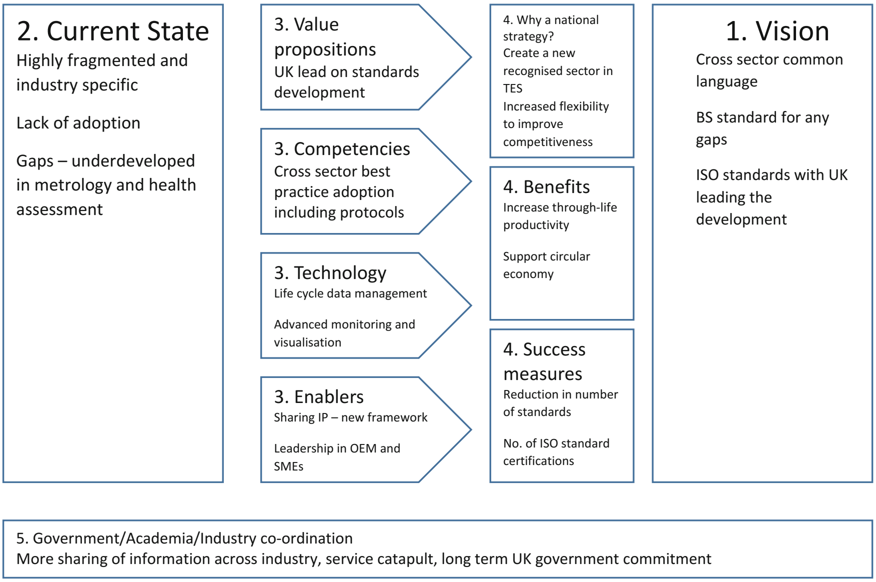


Fig. 3.5 Typical theme development template completed by a team for the Standards theme

3.3.2 Direct Workshop Outputs

The three workshops were given five themes initially and identified a further eight. The original five themes were technology and analytics, standards, supply chain, skills and collaborative mechanisms. The first four were a direct feed from the original white paper [1] and the fifth was the result of discussions at the National Strategy Steering Committee. The additional eight were culture and communications (including why TES), export development, innovation, market knowledge, design innovation, institutional leadership, productivity and finance. For each of these themes a number of teams had developed landscapes on the pro forma shown in Fig. 3.4. Figure 3.5 shows a typical output for the standards team generated by one team at the third regional workshop.

3.4 The Synthesis of the Workshop Outputs into a National Strategy

The three workshops produced twenty seven separate maps over the thirteen themes. The themes were consolidated into four key areas, global needs and opportunities (trends and drivers), the TES value proposition, TES capability development and enablers. For each of these classes of themes key arguments for why a TES strategy is needed were identified. This was supported by also listing the benefits and critical success factors that a TES strategy would engender. Finally a future vision was outlined and a list of required strategic actions developed.

In parallel with this work a second report had been commissioned on the market [2] and findings from this were incorporated into the overall strategy document [3]. The final strategy document [3] was developed and published in July 2016. All supporting documents are available to download from the TES Centre web site.

The strategy development process and its outputs highlighted three priority areas for action:

1. The creation of a cross sector UK National TES centre of excellence/Council with industry lead to:
    - Develop and promote 'TES thinking', broadening engagement across industrial sectors, sharing knowledge to develop supply chain capability and influencing sector strategies to encompass TES
    - Nurture innovation and improvement and associated investment and measure performance against the strategic KPIs
    - Provide a support network
    - Drive the TES educational agenda (including a TES module in every STEM course)
    - Direct the focus of research funding
    - Coordinate the development of future technologies to ensure the 'TES dimension' is not
    - Overlooked
    - Integrate government, industry and academia with the global scene
    - Drive the development of formal standards.
2. Government departments should change their procurement policy by putting emphasis on through-life costs.
3. A new government policy framework for IPR sharing should be developed, with open book services and a technology framework which covers IP law and improves protection.

Acknowledgements The authors wish to acknowledge the support of the EPSRC Centre for Innovative Manufacturing in Through-life Engineering Services in supporting the development of this strategy.

References

1. Tasker P (2015) Making things work engineering for life—developing a strategic vision. EPSRC centre for innovative manufacturing in through-life engineering services. http://www.through-life-engineering-services.org/downloads/27302_White_paper_amended.pdf. Accessed 23 Aug 2016

2. Hughes A, Hughes J (2016) Through-life engineering services (TES): market and data review. EPSRC centre for innovative manufacturing in through-life engineering services. http://www.through-life-engineering-services.org/downloads/Market_and_Data_Review.pdf. Accessed 23 Aug 2016
3. EPSRC Centre for Innovative Manufacturing in Through-life Engineering Services (2016) A national strategy for engineering services: delivering UK economic growth by making things work better for longer. http://www.through-life-engineering-services.org/downloads/Engineering_Services_Strategy_new.pdf. Accessed 23 Aug 2016

Part II

Through-life Engineering Services and Design

Chapter 4
Warranty Driven Design—An Automotive Case Study

Louis Redding

Abstract Much is to be found within the literature relative to Product Service Systems and the process of servitization. As companies adopt these philosophies the concept of advanced service provision and availability contracting continues to evolve and to be adopted. Underpinning all of this is the management and mitigation of risk to revenue streams due to the lack of availability of the product's design function resulting from failure or degradation. This has seen the emergence of Through-life Engineering Services and Integrated Vehicle Health Management which offer strategic solutions for the organisation seeking to mitigate risk and inform the design process. Whilst the majority of the literature focuses upon the aerospace and defence sectors little is known about the automotive sector. This chapter offers insight into the types of service data and the data sources used within a leading UK Automotive Manufacture. The means of obtaining the *voice of the customer* is also discussed. Finally an evolving 'straw man' TES informed system architecture is presented which will form the basis of future research and development by the author.

4.1 Introduction

Since their introduction to the literature by Mont et al. [1, 2], Product Service Systems (PSS) have continued to gain focus driven by academics, researchers, enlightened industrial practitioners, and organisations who seek to maintain and develop their competitive position. This growing understanding of the concept,

---

L. Redding (✉)
Centre for Innovative Manufacturing in Through-Life Engineering Services,
Cranfield University, Cranfield, UK
e-mail: louis.redding@cranfield.ac.uk

L. Redding
Jaguar Landrover Plc, Gaydon, Warwickshire, UK

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_4

frameworks, models, benefits and limitations are enabling practitioners within the manufacturing sector to seek new business models which enable them to ultimately offer whole life product/service solutions. The competitive pressures arising from the global market and in particular the emerging industrial nations (e.g. China, India, Brazil, etc.) together with increasing macro-economic, fiscal, environmental and legislative pressures have resulted in UK based organisations no longer being able to compete on the traditional cost, time, and quality alone [3, 4]. Response to these pressures has seen the adoption of differing levels of service provision in support of products by manufacturers as they move from being a pure manufacturer of products to the other end of Tukker's 'spectrum' [5, 6], that of a total service provider. The process of '*servitization*' has attracted a great deal of focus as manufacturers have sought to obtain strategic and competitive advantage. *Servitization* is seen as the process by which the organisation moves from a position of pure manufacturer of products to that of ever increasing integrated service solutions driven by the requirement to maintain and improve benchmarked competitive advantage [7–10].

The body of literature to date which has sought to investigate and develop the links between PSS, Servitization, and TES has tended to focus predominantly upon the Defence and Aerospace sectors with Rail Transport, Marine, and Energy starting to gain traction in this arena. However, contributions which relate directly to the automotive sector are proving to be scarce in number with no automotive manufacture offering near closed loop condition based management systems similar to those we see emerging within the civilian aero-engine manufacturing sector.

The automotive sector traditionally provides service support by way of extended warranty which is delivered by a complex structure of direct and franchised dealership networks. Whilst there are a plethora of data sources relating to vehicle performance and degradation in the field, supported by design and styling feedback, these data are often fragmented and difficult to trend. Typically, the diagnostics of degradation is achieved by directly linking the engine management system to a diagnostic system within the service facility (i.e. the approved service dealership) to achieve coded feedback from the sensors within the vehicle. The acquisition of such error codes and the corrective actions are the easy part of service support. Far harder is the collection, analysis, and reporting of the qualitative data that is the voice of the customer. This data set offers a very rich source of product information which has a profound impact upon the success of the OEM within the marketplace.

This chapter introduces the reader to these data sources and then offers a 'strawman' product development architecture which is based upon both the *voice of the customer* and the harvesting of '*hard*' fault data for consideration. It is hoped that by doing so, this chapter will help initiate further work within the area of TES enabled warranty systems and the development of system architectures that can harvest the intelligence provided by the data to further inform design.

4.2 Exploring the Link Between Warranty and Through-Life Engineering Services

The term ‘warranty’ can and does have several meanings, identities, content and context and varies depending upon the legislative nature of a given contact and the jurisdiction in which the contract was made. Whilst the literature is vast when seeking to identify and clarify the question “What is Warranty?” the contributions that are to be found are mainly within the Legal and Commercial Literature and add little value to this text. However, the subject of warranty definition, content, and context within the business contract space is worthy of a separate study and analysis, particularly when applied to through-life solutions enabled with TES. However, when seeking to identify a definition and understanding for the purpose of this work the author offers the following adopted definition for warranty. Namely:

> A written guarantee, issued to the purchaser of an article by its manufacturer, promising to repair or replace it if necessary within a specified period of time [Oxford English Dictionary:2016]

When considering this definition we find that Hogg [11] expands upon this. He states that a warranty is a guarantee or promise which offers assurance to the purchaser or operator of a product and that “specific facts or conditions are true…. [and/or]…will happen”. In the focus of this work this is taken to mean that the product will operate and deliver its ‘*design function*’ within specified and agreed limits of operation supported by an agreed service function and delivery system.

It follows that if the product manufacturer chooses to warrant the design function for a given time, or indeed through-out the usable lifecycle of the product, and also seeks to gain revenue through the use of availability contracting (e.g. Rolls Royce, Boeing, Pratt and Whitney, Bombardier etc.), then the manufacture would seek to mitigate the risk to the revenue stream due to product/system degradation or failure. To the customer or operator of the product, the warranty is akin to an insurance policy which protects the same from loss of the design function.

Whilst the literature focuses generally upon high value complex engineering products very little is written relating to the automotive sector when viewed through both the TES and Warranty lenses. Indeed, whilst conducting a literature search through ‘Scopus’ and ‘Web of Science’ which sought to link the subject matter relative to ‘TES & WARRANTY & AUTOMOTIVE’ the search returned *Zero* results. The need for further understanding and development of the links, content and context relative to TES and Warranty within the automotive sector is therefore plain to see.

Whilst such solutions give peace of mind to the private owner of motor vehicles and commercial operators, for the automotive OEM the provision of warranty services is essential to the business model. When reading the publication by the Telegraph Newspaper after an analysis warranty provision in the automotive sector by Foxall [12] we see….

> What these products tend to have in common……is their basic remit to retrain your business. That is often achieved by restrictive clauses in the policy terms and conditions which state that you must have servicing and repairs carried out at the supplying garage or dealership network. [12]

The automotive OEM however is exposed to many signals from warranty data and 3rd party survey data which can impact both positively and negatively upon the business. The issue with service data is that it is always time lagging. It relies upon the owner to present the vehicle for service within the time limits defined within the warranty. There is then further delay in receiving the intelligence relating to claims data as it is presented to the OEM via the dealership networks and approved franchises. In addition, whilst the 'hard' faults yield data which are quantitative in nature and therefore easy to trend, a significant amount of data relates to 'soft' faults and styling issues which are qualitative in nature and have poor customer verbatims in support of the claim. This results in time consuming and subjective analysis which can have significant impact upon the business.

When seeking to understand the possible causes for warranty claims Wu [13] suggests that these can be grouped into four categories (Fig. 4.1)

In reviewing the four groupings it can be seen that not all of the those identified relate directly to the product as one is customer centric and the other a reflection of the service provision. As will be seen in the next section of this chapter, this distinction is reflected in several 3rd party market surveys which seek to harvest and publish the customer experience relative to their vehicles. These surveys form a series benchmarked studies which cover the appeal, reliability and dependability of their cars, busses and trucks, the publication of which has a direct influence upon purchasing patterns and therefore the financial performance of each automotive brand and marque.

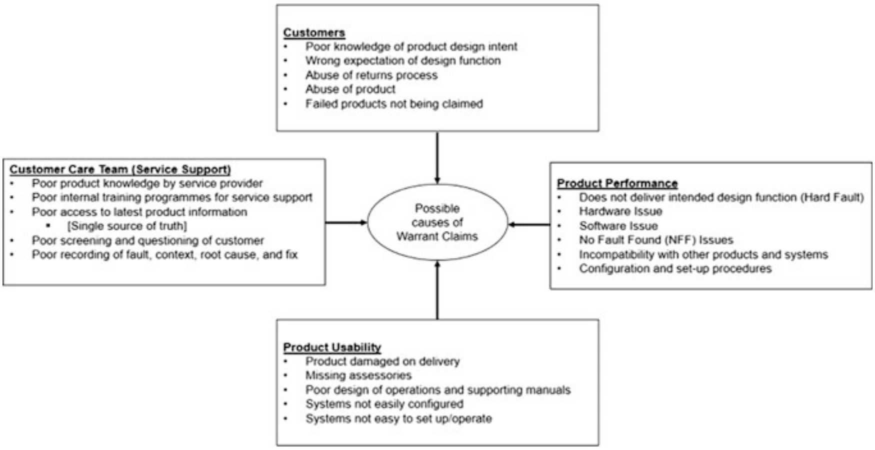


Fig. 4.1 Possible causes for warranty claims [13]

4.3 The 'Voice' of the Customer—An Automotive Example

When seeking to provide TES support for the product in the field it is imperative that one understands the performance of that product in a timely manner. This is normally achieved by the application of either condition based monitoring systems to the product which then transmit data in either real time (IVHM generic) or through physical download activities within the service centre. This information is performance data or engine management codes and can also be obtained by the indirect reporting of degradation and failure through the harvesting the Voice of the Customer. This section will focus upon the latter. The acquisition of such data is essential when seeking to mitigate risk to business revenue, the manufacturer competing through such support offerings as extended warranty which is enabled by TES solutions, be they either 'intermediate' or 'advanced' levels of service (ref Baines).

When wishing to obtain this data within the sector several sources are available to the design and quality functions in order for them to continue to improve that design and to reduce the number of 'hard' and 'soft' failure modes (Fig. 4.2).

The following subsections will introduce the reader to these data sources, most of which are publicly available albeit some are subscription based.

4.3.1 JD Power

JD Power is an international data source for obtaining the voice of the customer relative to products and services. The organisation conducts full market surveys,

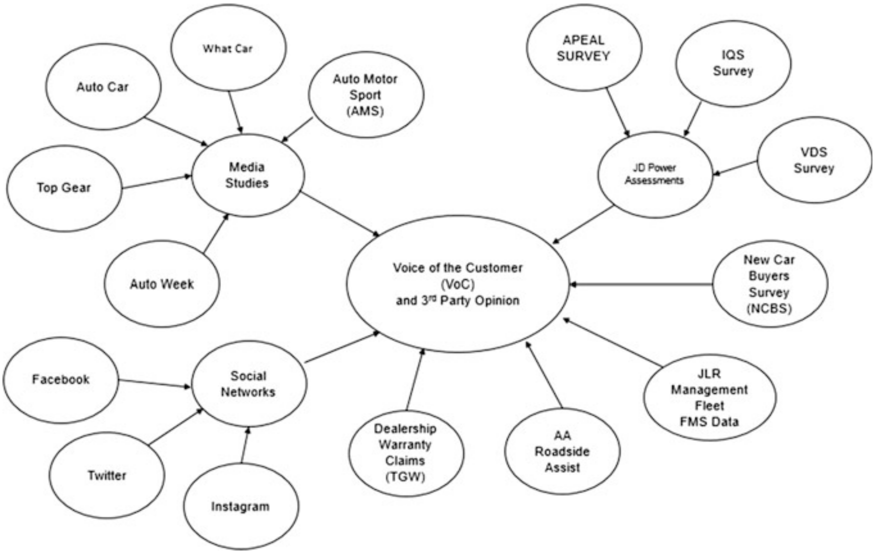


Fig. 4.2 Means of acquiring the Voice of the Customer (VoC) within Jaguar Landrover

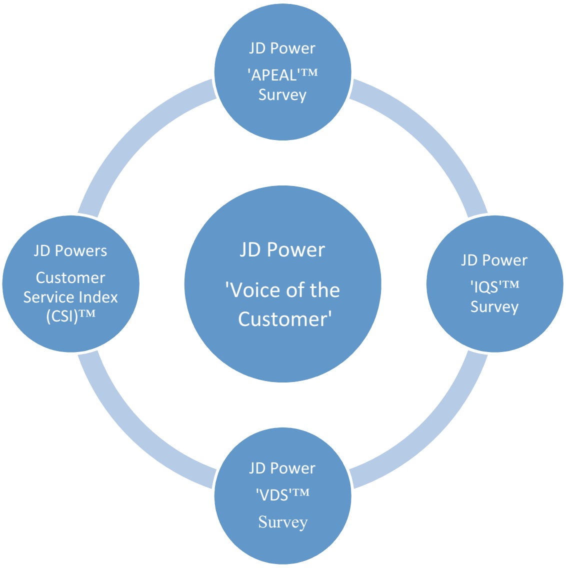
delivers analysis, and publishes benchmark studies which are used by all the major automotive OEM's in assessing the appeal of vehicle design offerings, quality and reliability. It does this through four lenses which seek to focus upon the (i) Level of customer service offered by the OEM (or its franchises) in support of its products, (ii) The dependability of the product, (iii) The *actual* and *perceived* quality of the product, and (iv) The performance and design of the product. The findings of analysis is published in four benchmarked data sets (Fig. 4.3)

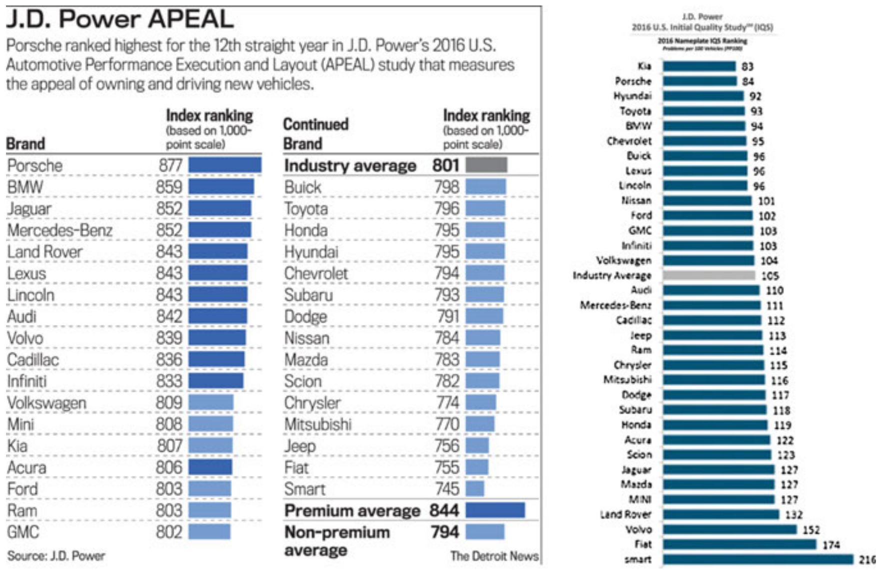
This data set is a key driver for the design of future vehicles as when viewed collectively it offers the OEM direct market intelligence into the benchmarked appeal of its design, styling and perceived quality and finish. Whilst these are subjective in nature and vary with automotive fashion and trends, the benchmarked studies prove invaluable when seeking to design and launch the next vehicle programme and engineer the next generation of service and warranty offerings.

The Automotive Performance, Execution and Layout (APEAL™) Survey (often referred to as '*Things Gone Right*') publishes benchmarked analysis of customer satisfaction relative to the design, styling, performance, content and internal layout (trim etc.) of vehicles during the first 90 days of ownership (Fig. 4.4a). Typically the study is based upon eight categories of performance. Namely:

- Engine and Transmission,
- Ride, Handling and Braking,
- Comfort and Convenience,
- Seats,
- Cockpit and Instrument Panel,
- Heating, Ventilation and Cooling,
- Sound System (Info-tainment),
- Styling and Exterior.

Fig. 4.3 Sources for the '*voice of the customer*'—JD Power





J.D. Power APEAL

Porsche ranked highest for the 12th straight year in J.D. Power's 2016 U.S. Automotive Performance Execution and Layout (APEAL) study that measures the appeal of owning and driving new vehicles.

Brand	Index ranking (based on 1,000-point scale)	Continued Brand	Index ranking (based on 1,000-point scale)
Porsche	877	**Industry average**	**801**
BMW	859	Buick	798
Jaguar	852	Toyota	796
Mercedes-Benz	852	Honda	795
Land Rover	843	Hyundai	795
Lexus	843	Chevrolet	794
Lincoln	843	Subaru	793
Audi	842	Dodge	791
Volvo	839	Nissan	784
Cadillac	836	Mazda	783
Infiniti	833	Scion	782
Volkswagen	809	Chrysler	774
Mini	808	Mitsubishi	770
Kia	807	Jeep	756
Acura	806	Fiat	755
Ford	803	Smart	745
Ram	803	**Premium average**	**844**
GMC	802	**Non-premium average**	**794**

Source: J.D. Power

The Detroit News

Fig. 4.4 **a** J.D.Power APEAl 2016 Survey Rankings. **b** J.D.Power 2016 IQS Survey Study Rankings

The survey harvests customer responses relative to 90 vehicle attributes with JD Power stating that its 2016 report is based upon a sample size of 80,000 vehicles. The reports obtained by the automotive design function from this analysis covers both ‘hard’ and ‘soft’ faults.

The Initial Quality Survey (IQS™) (often referred to as ‘*Things Gone Wrong*’) reports upon the quality issues (both hard and soft faults) that have been experienced by customers within the first 90 days of ownership (Fig. 4.4b). The findings are reported as a benchmarked study based upon 200 individual vehicle parameters which cover the quality and mechanical design of such areas as:

- Powertrain
- Transmission
- Body and Interior build quality and finish
- Features
- Accessories

Whilst the findings of the IQS survey offer essential data when seeking to monitor and improve quality and design, and form one of the key ‘in-service’ feedback mechanisms for the designer it is helpful to have a long term perspective of vehicle performance.

The Vehicle Dependability Study (VDS) goes part way to solving this issue by examining and reporting upon quality and reliability problems that are experienced when the vehicle is greater than three years old and is still owned by the original purchaser. The survey and analysis reports upon the reliability and dependability of the product (vehicle) during the previous twelve months of ownership (or operation). The features covered are generally aligned to those within the IQS survey with the metric being the number of problems for a given system measured in problems per 100 vehicles (PP100). This figure is used to ascertain the dependability of the vehicle relative to a given system. Whilst a given automotive OEM has a rich data source obtained from its authorised dealer network and warranty prevision through its franchises, the data that is provided within the VDS study gives additional benchmarked studies for key engineering systems (engine, transmission, braking etc.).

When reviewing the data for 2016 the author suggests that it is interesting to note that the majority of dependability issues are in the electrical, and electromechanical systems of the vehicle and are centred on design related issues. When reviewing part of JD Powers summary it is observed that:

> Among the study’s key findings, perhaps most notable is that the overall industry average is 152 PP100 this year….[2016], compared with 147 PP100 in the 2015 study. Seven of the top 10 problems are design-related. Design-related problems account for 39% of problems reported in the 2016 study (60 PP100), an increase of 2 percentage points from 2015. [http://www.jdpower.com/cars/articles/jd-power-studies/2016-vehicle-dependability-study-technology-woes]

This is seen by the author as being very significant when one views the current drive towards autonomous self-drive vehicles and the associated system design and

No Fault Found (NFF) issues that are experienced within such systems which impact upon reliability and assured resolution of system degradation and failure.

Renne Stevens, the Vice President of Automotive at JD Power recently (2016) states that:

> Usability problems that customers reported during their first 90 days of ownership are still bothering them three years later in ever-higher numbers. At the same time, the penetration of these features has increased year over year…………..If you think about the technology problems from the study in the context of conversations around autonomous vehicles, the industry clearly has more work to do to secure the trust of consumers. …….Right now, if consumers can't rely on their vehicle to connect to their smartphone, or have faith that their navigation system will route them to their destination, they're certainly not yet ready to trust that autonomous technology will keep their vehicle out of the ditch. [http://www.jdpower.com/cars/articles/jd-power-studies/2016-vehicle-dependability-study-technology-woes]

Whilst the challenges that exist relating to the next evolutionary change relative to the design, operation and ownership of automobiles are significant and will be covered by further publications by both the author and no doubt other researchers within the TES community, this focus is worthy of a significant text within its own right and is not covered in this chapter. Needless to say however that the author is of the opinion that the solution of these challenges will have a major impact upon all aspects of the automotive sector, integrated transport systems, and the macro economy.

Finally, **The Customer Services Index (CSI)™** surveys and reports upon the customer's (or operator's) satisfaction with aftersales services support and through-life of ownership service and support where appropriate which is usually offered by the dealership networks. Here customer opinions are sought and reported of recent dealership service support for both warranty and customer paid work. The measures reported are:

- The service initiation experience
- The vehicle collection experience
- The standard of the service facility
- The perception of the service quality
- The interaction with the service provider (interpersonal relationship etc.)

The metric however does break down the reporting into Premium Brands and Volume Brands (Fig. 4.5).

In separating these two metrics into either '*Premium*' or '*Volume*' Producers one assumes that behind this distinction lies differing brand distinctive service solutions as implied by Baines (see Sect. 4.2). Here the author assumes that the OEM in conjunction with its dealership network offers both 'Base' and 'Intermediate' service provision in support of the product. However, with Premium Brand Vehicles (e.g. Rolls Royce, Bentley, Porche, High end BMW etc.) the service support will be far more intimate during routine warranty and servicing, and MRO where things go wrong. For high end vehicles one sees such support as that offered by Rolls Royce Automotive who provide a four year comprehensive package which is regardless of the mileage driven and which includes all servicing, repairs, maintenance, and road

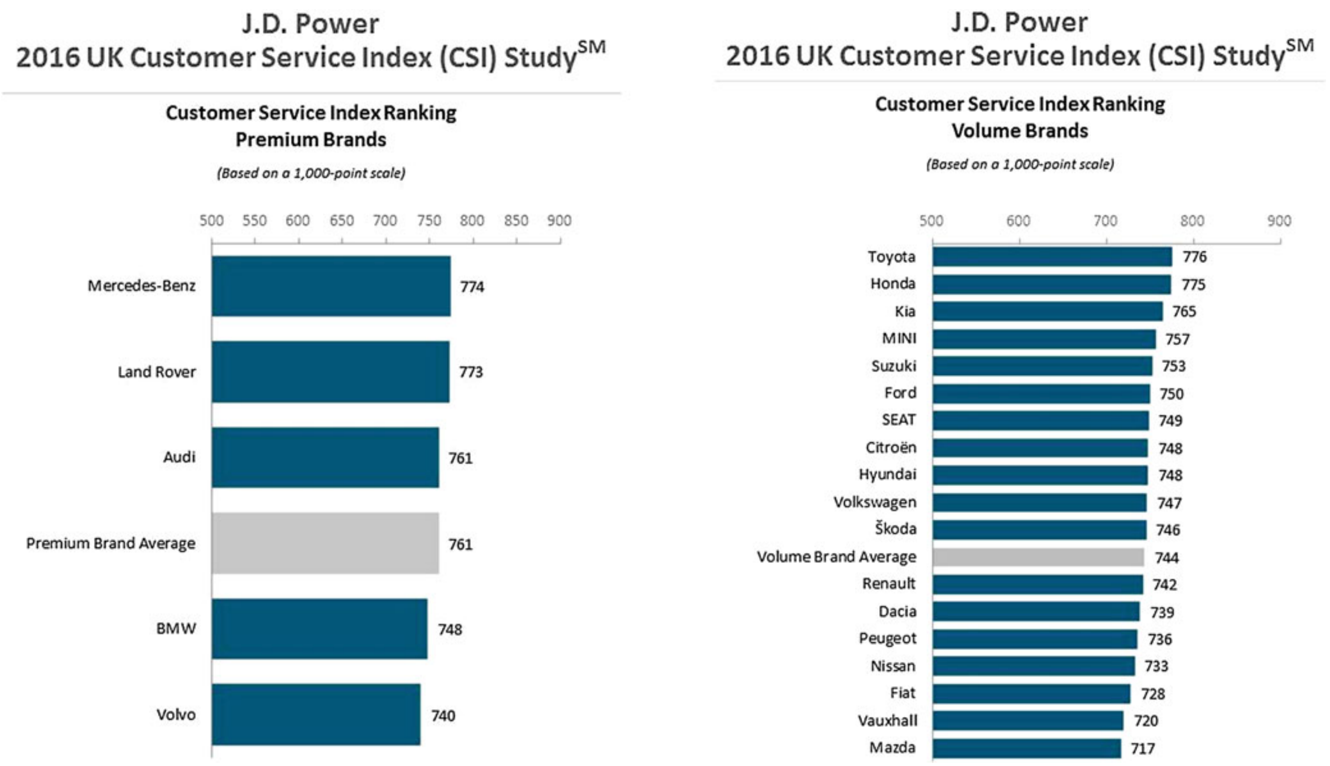


Fig. 4.5 Customer Service Index (CSI) for Premium and Volume Brands Source: JD Power CSI Survey (2016)

side assistance by a dedicated Rolls Royce Service Operation. In addition, when reviewing Rolls Royce's advertising literature one reads:

> To make servicing faster and more effective, every new Rolls-Royce features Condition Based Servicing. This means sensors in your Rolls-Royce actively monitor important components such as engine oil, brake pads and filters. If you've signed a 'Get Connected' form, this information will be transmitted to your dealership so that services can be scheduled to meet the precise needs of your vehicle, at a time most suited to you. [https://www.rolls-roycemotorcars.com/en-GB/ownership.html]

The company also offer the Rolls Royce Enhanced Ownership Programme™. This uses further innovative technology which is IVHM and CBM generic to ensure that only the highest level of service support is offered in the support of the product.

> The programme has two features. Rolls-Royce TeleServices is a unique offering that delivers a seamless maintenance experience, while Rolls-Royce Assist offers a swift and effective response in times of a vehicle emergency. [https://www.rolls-roycemotorcars.com/en-GB/ownership.html]

The CSI report is an essential data source therefore when seeking to scope and understand the competitive space of the organisation relative to the service delivery system. It helps to inform the service support strategy by offering benchmarked findings relative to the service content and customer experience when seeking to design a TES enabled service support and delivery system. In the highly competive world of the automotive sector it can be seen without a large amount of imagination that to be a market leader, an understanding relative to the threshold of expected quality and reliability of the product is essential. This is becoming evermore important and the ability to apply TES generic technical and business solutions in order to maximise these attributes is of paramount importance in the drive to compete (Table 4.1).

The JD Power dataset and suite of benchmarked reports are an essential resource to all automotive OEM's who seek to enhance their Product Creation and Delivery Systems (PCDS) whilst meeting customer expectation relative to the actual and perceived quality of both the product and the service delivery system.

4.3.2 New Car Buyers Survey (NCBS)

This European study is conducted by the NCBS and seeks to harvest by the 'rational and emotional' opinions of customers post the purchase of their vehicle. This study provides subscribers with a very detailed analysis of these opinions by market segment, marque and model, and country of sale. The data illustrates the complexity of rational and emotional factors which have to be taken into consideration when designing a new vehicle programme and supporting service delivery system. The links between varying aspects of expectation are very complex and not always clear.

However, by study of the Metadata by complex algorithms, various benchmarked market trends emerge which aid the product content and styling supported by service

Table 4.1 Rolls Royce Automotive Service Provision

Rolls Royce Tele Services™	Rolls Royce Assist™
“Rolls-Royce TeleServices™ is an advanced system that allows your car to transmit important service data to your authorised dealer with exceptional accuracy. Whether your Rolls-Royce needs an oil change or a battery charge, your dealer will be alerted and can quickly arrange for the necessary work to be carried out at a time that suits you”	“Rolls-Royce Assist is an unparalleled system that connects’ your motor car with an emergency services call centre. The system is activated in two ways:
	In the event of activation of the airbag or crash sensors, the vehicle will automatically trigger a call to the neerest emergency services call centre. Already aware of your location, the emergency services will then attempt to make contact with the driver or passenger and dispatch the emergency services if no response is received.
	If you find yourself in an emergency situation, you can manually activate Rolls-Royce Assistsimply by pressing the SOS button. This will correct you directly to the emergency services call centre, and enable “you to explain your situation to them.
	Rolls-Royce Assist is designed to offer you reassurance, so that nothing detracts from your pleasure while driving your motor car”

Source: https://www.rolls-roycemotorcars.com/en-GB/ownership.html

support which is aligned to the expectation (and fashion) of the market. It is also important to note that the distinction between vehicle marques relative to product offering, performance, and reliability and dependability for a given class of vehicle as diminished as all design (and customer expectation) trends towards a norm. For any given class of vehicle the range of product offering is becoming ever-more convergent. This makes the importance of the ‘emotional’ factors increasingly significant. The ongoing development of zero defect initiatives and customer delightedness is central to design, the purchase decision, and through-life vehicle support (warranty).

4.3.3 AA Roadside Assist and Dealership Warranty Claims

Two additional sources of product reliability data are those registered by the emergency breakdown services (e.g. Automobile Association (AA) and the Royal Automobile Club (RAC)) and the direct warranty data which is available after service by the dealerships. The OEM’s promote the purchase of service contracts as symbols of warranty. In Majeske et al’s [14] study into the evaluation of product and process design changes with warranty data they state:

> In the calendar year 1992 the big three (Ford, General Motors, and Chrysler) spent a combined $9.2 billion on direct payments to authorized service centers for repairs covered by warranty [14]

Whilst this data is somewhat dated the author can advise from working within an automotive OEM that Jaguar Landrover spent GDP 265 Million on Warranty Claims in 2015/6 which equates to the design and launch of a new programme upon an existing vehicle platform for its Body Engineering Function. It is therefore seen that this expenditure is very significant when reviewing the balance sheet and Profit/loss accounts for automotive OEM's.

These figures also provide an indirect measure of customer satisfaction. Indeed, this information is also available directly to the public as the aftermarket cover provider Warranty Direct publishes a 'Reliability Index' detailing and benchmarking brands and marques. The guide informs the reader of the probability of vehicle and component failure by use of it's index.

The common approach to modelling these failure rates is to use the following formula which expresses the rate using the number of repairs per 1000 vehicles.

$$R(t) = \sum_{i=0}^{t} \frac{\text{Claims}\,(i) \times 1000}{N(t)} \tag{4.1}$$

Where:

R(t) = repairs per 1000 vehicles
T = months in service
N(t) = number of claims observed in month t by N number of vehicles

However, using this method can overestimate the warranty figures if the same vehicle presents a warranty claim several times in a given period. There is always a risk of double counting if due diligence in the analysis is not carried out.

Whilst these figures are commonly used within the sector from the sources specified, detailed figures which relate to reimbursement of warranty costs to the dealerships exist upon OEM bespoke quality systems and offer a rich vein of intelligence relative to failure and degradation rates which can greatly inform the design function of the manufacturer.

4.3.4 OEM Fleet Management Systems (FMS)

Most large automotive companies offer key employees access to their products by way of Corporate Agreements for the use company cars which for managerial grades, forms part of their remuneration package. Such systems offer an easy to use platform from which to manage all aspects of the organisations vehicular assets. For Jaguar Landrover Plc the system offers a centrally managed solution which facilitates a transparent access to all aspects of vehicle assignment, use and performance, service, throughout the whole life-cycle of the asset. That is from cradle to the grave.

Within JLR the system typically monitors and controls 3800 vehicles in 83 geographic regions across five functions of the business. This represents a capital value of approximately USD 300,000,000 [Source: JLR Internal Reporting]. This system has and continues to yield significant advantages to the organisation since its inception.

> To date FMS has logged hard savings in the region of $6.5 million pa through inventory control, warranty reduction and more efficient use of our assets [Source: JLR Internal Reporting]

The use of this fleet data assists in the feedback of TES generic and usage data which in turn aids design of future programmes, forward model quality, and reliability. Additionally the data harvested also aids and informs future design by identifying 'soft' issues which include Perceived Quality relative to internal/external trim levels, comfort, NGV issues (Noise, Vibration and Heat) etc. This in turn helps to drive down warranty cost by:

i. "….the improved Driver Appraisals and Issue Search …[functions]….ha[ve] resulted in Issue Reviewers being able to search for a problem not only within a small fleet ….but across the entire range of vehicles on the system"
ii. "…as the Own Use vehicle progresses through its active life it collects a detailed history of all of its activities—this enables an informed decision to made at a later date on whether to rework the vehicle (at significant cost) and sell it with a warranty or to scrap it."

Whilst this system monitors the issued cars for the fleet, it is also used for pre-production vehicles which are released to the management of the organisation. The rationale for this is to get real performance data of the cars from 'informed' drivers. This tests from pre-volume production use yields significant quality data and tends to be harsher than the results found from the general public. It therefore offers a significant opportunity to resolve any legacy issues prior to volume launch.

4.3.5 Media Systems

The automotive media (Television and Print) offer a plethora of expert and consumer opinion ranging from professional test drive reporting, customer surveys, buying trends, 3rd Party Customer 'Watchdog' Publications, and general interest news stories. Typically the main sources of such information are AutoCar™, WhatCar™, Auto MotorSport™, Auto Week™, and TopGear™. These data sources have great significance as they can also inform and drive customer opinion, whilst in the case TV Programming, offering entertainment also. The latter's ability to both entertain and inform is very powerful as it can directly influence the Brand Positioning.

4.3.6 Social Networks

This method of performance and quality reporting of both the product and service is becoming ever-more significant and has the potential to greatly influence (either positively or negatively) the Brand of the product. Typically such platforms as Facebook and Instagram (Twitter) are seen by the consumer as a means of expressing opinion and dissatisfaction relative to product quality and performance, and level of service. What makes these platforms so important to the manufacturer is the potential for postings to go viral in near 'real time' and with a world reach. This has led to some automotive OEM's having departments that monitor the transmissions on these platforms and where necessary execute real time solutions to the customer so as to avoid adverse damage to the brand offering. This is not just a marketing initiative however as the findings are relayed directly back to the organisation's design and quality functions who are tasked with providing both the containment and the long term design solution.

This section as served to illustrate some of the mechanisms by which automotive OEM's obtain the 'voice of the customer'. It has highlighted the amount of data that is readily available for use which relates to both engineering function and subjective opinions and preferences. It can be clearly seen that the role of TES becomes increasingly important if a whole life customer support system is to be provided and that the issues raised include both 'hard' faults and 'soft' issues which can equally damage the manufacturer if not addressed in a timely manner. It is also no longer sufficient to rectify an issue. It is of equal importance that the manufacturer's reaction to a given issue is seen to be timely and effective if the organisation is not to be damaged by the expression of negative opinion that can reach a critical mass through the use of social media.

4.4 Development of a TES System Architecture to Inform Design

It has been seen that designers within the automotive sector have a plethora of data sources from which to harvest information and knowledge with reference to their product and design performance. These data relate not only to the hard failure modes which typically relate to loss or degradation of the design function and therefore the ability to use, but also the softer elements of the voice of the customer which relates to opinions of the product and service support offering. Where there is negative feedback relating to the support offering (i.e. The Customer Services Index) it is hoped that this would trigger investigation as to the reason for such criticism. Typically this subsequent investigation holds the potential for valuable feedback relative to design for service. Whist a significant amount of data relates to customer *front line* service dissatisfaction from the service transaction, valuable data is also recovered from the dealerships on the ease for which the service action can be carried out. In other words, is the design service friendly for the engineer/technician who is tasked with carrying out these activities?

When seeking to evaluate the current performance of a vehicle system, product, or styling feature offering, the designer has an architecture (formal or informal, known or unknown) which holds significant amounts of data and metadata from which to call upon (Fig. 4.6). Whilst this architecture is seldom fully defined within the sector OEM's and First Tier Supply organisations, and whilst there are no contributions to the literature which seek to fully map these data sources and draw the links between Warranty, TES, and design, the architectural map exists. It is available if only the designer knows where to look and how to harvest the information readily available. It is only by being able to undertake this approach that robust designs and service solutions are developed. The architecture illustrates how 'in service' warranty data and TES generic data can, and does inform the design of cars, trucks and buses.

When reviewing Fig. 4.6 one sees that there are numerous data sources, most of which have been introduced and discussed in Sect. 4.3. It should be noted that whilst this paper has focused upon the use of the qualitative data that is available in the sources listed, there are also available other sources of explicit and tacit knowledge that resides within the organisation from which the designer can call upon. Most of which is, or can be codified based upon the scientific method.

From these data, information is generated through the use of various engineering and quality techniques and by the use of bespoke platforms, aligned reporting is generated. Within Jaguar Landrover the following systems are widely used:

- PAWS—Integrated Issue/Project Management System
  - This system provides a standardised, globally accessible intranet tool for analysing and prioritising warranty claims and TGWs with effective project-based problem management
- AIMS
  - Automated Issues Management System
- Catia
  - CAD Drafting and Modelling System

JLR systems typically harvest all of the data from the aforementioned data sources and imports them into their Integrated Issue and Project Management System. Individual issues are grouped into projects by vehicle line, market sold, trim level, component and then failure mode. Explicit hard fault data and the voice of the customer is recorded within the system and is available for recall by the designer who can choose by selection of pre-set filters to output the data as cross referenced data sets. At the start of the design process these reports are used to update the issues log and quality history files prior to commencement of system and component design. In addition, it facilitates the strategic decision to fix or leave an

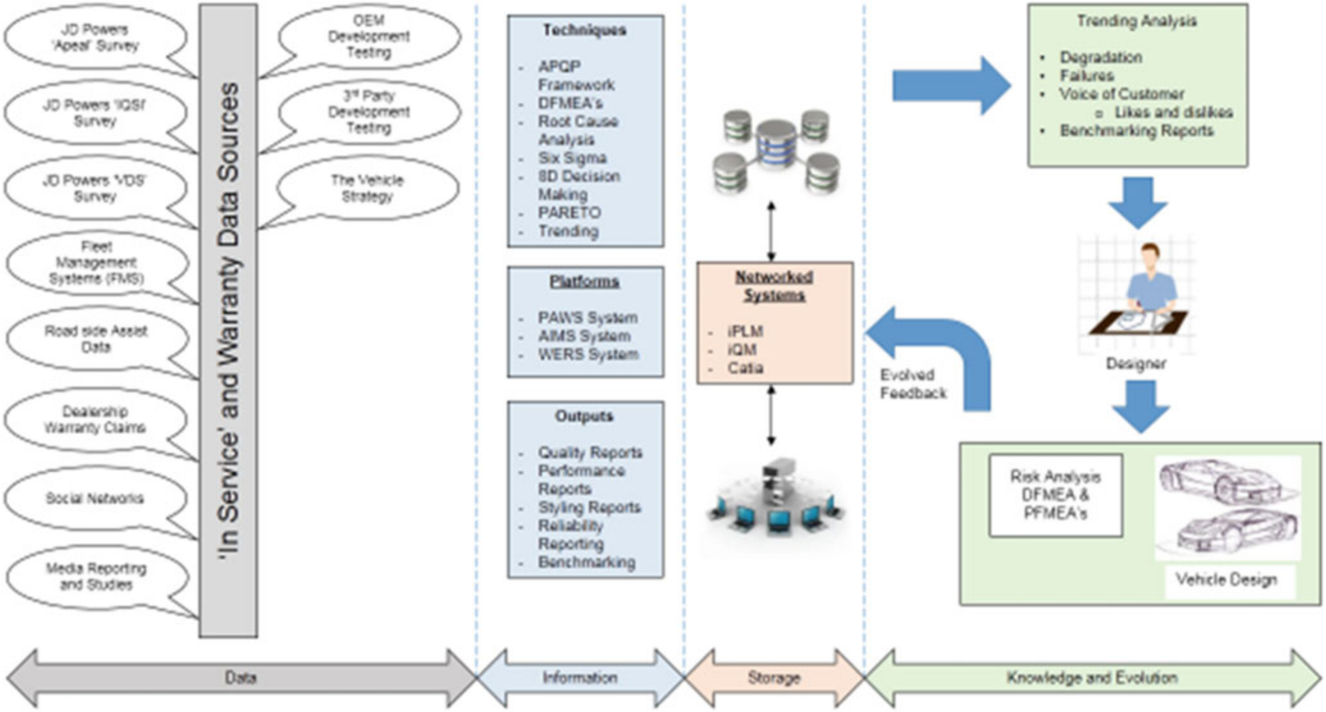


Fig. 4.6 A typical service and warranty data architecture available to the automotive design function

issue subject to the volume of customer verbatim and the business case for any rectification actions.

The input, storage, and outputs relative to these activities are driven by central networks within an iPLM (Product Life-cycle Management), and iQM (Quality Management) framework. By so doing the organisation is driving towards a 'single source of truth' for its product design, manufacturing, and technical service functions which is available within its global operation.

The final part of the figure shows the role of the designer. As the reader places his or herself within the position of the designer it is seen that at the start of the design activity the trending studies relative to system and component degradation and failure are studied. This is in addition to the benchmarked studies that report out upon the driver and service experience, the likes and dislikes relative to styling and functionality of trim levels and ergonomics. Typically the designer will then have to ensure that this data aligns to the proposed vehicle strategy (specification) as defined by the organisation's board. Then the work relating to risk analysis (i.e. component and system degradation and failure) can start as robust DFMEA's, PFMEA's, and Measurement and Control Plans are defined that comply with the accredited APQP (Advanced Product Quality Planning) process. It is only once all these activities are successfully completed can the task of actual design, prototyping, testing, and finally manufacture can actually start. It can be seen that the designer, supported by the quality function, actually sits upon a mountain which is constructed upon data and when in context, information. It is only when gaining an understanding of these concepts and their interactions does the designer gain true knowledge and the ability to generate service informed designs which will mitigate risk of design function, thus maximising availability for use and thus generate maximum revenue through transactional contracts and/or a through-life availability contract. The challenge remains and that is how to '*join up the dots*!'

4.5 Concluding Remarks

This chapter has introduced the reader to the data and its sources relative to TES and Warranty used within the automotive sector. Upon review of the literature it is seen that the majority of contributions relate to the aerospace, defence, rail, and marine sectors with very little existing which focusses upon the automotive sector. Of those contributions the literature discusses mainly bespoke technical issues relative to component and service degradation, cost modelling, self-healing and autonomy. The author would suggest that the automotive industry is more challenging when seeking to understand these concepts. The actual operator of the products covered in these sectors are generally highly trained and work within narrower operating practices. For example, the civilian airline pilot is very highly trained, operates within well-defined flight paths with constrained flight profiles. The train driver is also strictly confined to bespoke routing and acceleration and cruising speed restrictions which are monitored and externally controlled by advanced signalling and braking systems.

Within the automotive sector however, the parameters of driver behaviour are very broad with component life being greatly influenced by driver profile. Excessive acceleration, braking and poor clutch control all have significant impact upon service life. So does the driver's attitude to service intervals and MRO procedures and intervals. It is therefore significant that there are very few successful modelling initiatives if any that seek to optimise availability and align TES generic support. But things may be changing.

When reading the media and observing social behaviour we see such organisations as UBER emerging. Here, the linkage with the smart phone results in UBER's statement of "*Uber is the smartest way to get around. One tap and a car comes directly to you. Your driver knows exactly where to go*". The system is based on car sharing where the individual signs up to UBER and is then connected to others seeking the same journey who are willing to pay for the ride. This then makes the daily commute a revenue stream for the driver which in turn makes vehicle availability more significant. The significance with UBER is that it is promoting car sharing and as such it is not too much of an assumption to state that this could make the shift in views relative to routine car ownership. Why own the car which is a depreciating asset when you can buy the journey.

There is also the increasing incorporation of in vehicle intelligence. From passive sensor technology which gives audible signals if the car is too close to an external object, to satellite navigation guidance system which are slowly becoming closed loop we see the emergence and the evolution of autonomy. For the author, these are only foundation technologies from which the autonomous driverless car will emerge. There are current trials in Milton Keynes (UK-2015/6) of autonomous cars and this is also developing in parallel in the USA driven by the TESLA and its flamboyant CEO—Elon Musk.

In addition, there are currently several planning initiatives for a new 'Smart City' on the outskirts of London with a fully integrated transport system. Within the urban area it will be fully pedestrianized with only autonomous vehicles providing transport within the centre of the city.

The point of raising these initiatives here is obvious. Quite simply the future of volume production of the automobile could radically change as people start to think about the need for car ownership, its advantages and disadvantages, when for most, all they wish to do is to purchase the journey. For the operators of such automotive vehicles within this brave new world the reliability and availability for use within the integrated transport system will become of paramount importance to their ongoing revenue streams. If one accepts this evolution as being inevitable then the need for effective TES system architectures which also harvest warranty data and can therefore be used to directly inform design via closed loop systems becomes essential. The author suggests that society is moving towards a new frontier, the boundaries of which are measured only by our own will to embrace change which could be either transitional or transformational.

References

1. Mont O (2000) Product-service systems. J Eng Des 20(4):327–328. doi:10.1080/09544820903149271
2. Mont OK (2002) Clarifying the concept of product-service system. J Clean Prod 10(3):237–245. doi:10.1016/S0959-6526(01)00039-7
3. Treacy M, Wiersema F (1997) The discipline of market leaders: choose your customers, narrow your focus, dominate your market. Biogr Interdiscip Q 210. doi:Book Review
4. Redding LE (2012) A strategy formulation methodology for companies seeking to compete through IVHM enabled service delivery systems. Cranfield University, Cranfield, Bedfordshire, UK
5. Tukker A, Tischner U (2006) Product-services as a research field: past, present and future. Reflections from a decade of research. J Clean Prod 14(17):1552–1556. doi:10.1016/j.jclepro.2006.01.022
6. Tukker A, Tischner U (2004) New business for old europe. Suspronet 132–148. http://www.ncbi.nlm.nih.gov/pubmed/18048240
7. Vandermerwe S, Rada J (1988) Servitization of business: adding value by adding services. Eur Manag J 6(4):314–324. doi:10.1016/0263-2373(88)90033-3
8. Wise R, Baumgartner P (1999) Go downstream: the new profit imperative in manufacturing. Harvard Bus Rev 77(5):133–141.doi:[WB99]
9. Oliva R, Kallenberg R (2003) Managing the transition from products to services. Int J Service Ind Manag 14(2):160–172. doi:10.1108/09564230310474138
10. Neely A (2009) Exploring the financial consequences of the servitization of manufacturing. Oper Manag Res 1(2):103–118. doi:10.1007/s12063-009-0015-5
11. Hogg M (2011) Promises and contract law. Cambridge University Press, Cambridge, UK, p 48
12. Foxall J (2015) Extended car warranties—are they worth it? The telegraph. Feb 2015
13. Wu S (2012) Warranty data analysis: a review. Quality Reliab Eng Int 795–805. doi:10.1002/qre.1282
14. Majeske KD, Lynch-Caris T, Herrin G (1997) Evaluating product and process design changes with warranty data. Int J Prod Econ 50: 79–89. doi:10.1016/S0925-5273(97)00034-0

Chapter 5
Designing for Service in a Complex Product Service System-Civil Aerospace Gas Turbine Case Study

Andrew Harrison

Abstract Gas turbine engines used in commercial aviation are a complex engineering product for which the cost of maintenance over their full operational life can exceed the initial cost of manufacture by a factor of 3–4. With an increasing shift to contracting fixed price maintenance services at the point of sale of the engine, it has become an imperative that the cost of maintenance support is a prime design attribute. This chapter provides a case study on the process of designing for an optimal maintenance cost level from the earliest stages of conceptual design through to product entry into service. The nature of an optimal maintenance cost in terms of customer service value is also explored.

5.1 Case Study Context: The Rolls-Royce Aero Engine/TotalCare ® Service

Rolls-Royce is a global company with a portfolio of products and services predominantly in power generation[1] systems across a range of applications.

The majority of our products have a long lifespan of typically 25 years or more and many of our product lines will have an in-service span (1st production to final disposal) in excess of 50 years.

Rolls-Royce is widely credited with bringing the *Power by the Hour*® service contract to the civil aviation market with its TotalCare® service support contract, in which the engine operator contracts a fixed price suite[2] of support services on a $

---

[1] Rolls-Royce has 49,900 employees worldwide and is active in the Commercial Aerospace, Defence Aerospace, Commercial and Naval Marine and Energy sectors (as of 2016) [1].

[2] TotalCare® comprises a suite of engine support services including; proactive in-service engine management, overhaul shop maintenance and the collection and management of engine data.

A. Harrison (✉)
Engineering Associate Fellow (Life Cycle Engineering), Rolls-Royce Plc,
PO Box 31, Derby DE24 8BJ, UK
e-mail: andrew.harrison@rolls-royce.com

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_5

per flying hour basis. TotalCare® contracts can cover just a few years of operation up to the full operational life of the product. This contrasts with the traditional contracting model in which the operator commissions and pays for individual maintenance events as they are required (commonly known as Time and Material contracting).

With a significant proportion of engine sales and support contracts being contracted during the product design and development phase[3] there is a significant transfer of cost management risk from the operator to the supplier. Maintenance contract prices are often established a decade ahead of statistically significant service cost data being available.

The advantages of TotalCare® are the alignment of operator and engine supplier benefits and incentives. Under TotalCare® it is clearly in Rolls-Royces interests to design a product that not only meets the operational guarantees but also delivers a cost of support in line with the contracted rate. The return for Rolls-Royce is the stability of committed future service revenue streams.

It is this transfer of long term support cost risk to the supplier that brings the need to design for service into such sharp focus.

5.2 The Costs of Owning a Gas Turbine Aero Engine

The following case study is based around a gas turbine engine for the civil aviation market (visualize the engines hanging from the wings of the last commercial aircraft you travelled on). Whilst the examples may be specific to this application many of the features of the scenario are common to the majority of complex, high value, long life products.

One of the major difficulties with designing for service contract risk management is the sheer complexity of the cost structure and the number of degrees of freedom that impact its evaluation. The following sections describe in outline the key elements of the cost breakdown and some of the primary driving factors that influence them.

In general all costs associated with the ownership of an engine fall into one of two categories:

- Direct consumables—costs incurred as a consequence of normal operation and performance of the engine (fuel, oil).
- Loss of functionality costs—costs associated with the prevention or recovery from loss of engine functionality (due to degradation, damage, or other causes).

---

[3] Rolls-Royce announced further sales of Trent XWB engines with contracted TotalCare® services at the 2009 Paris air-show 4 years ahead of the expected entry into service date of the aircraft.

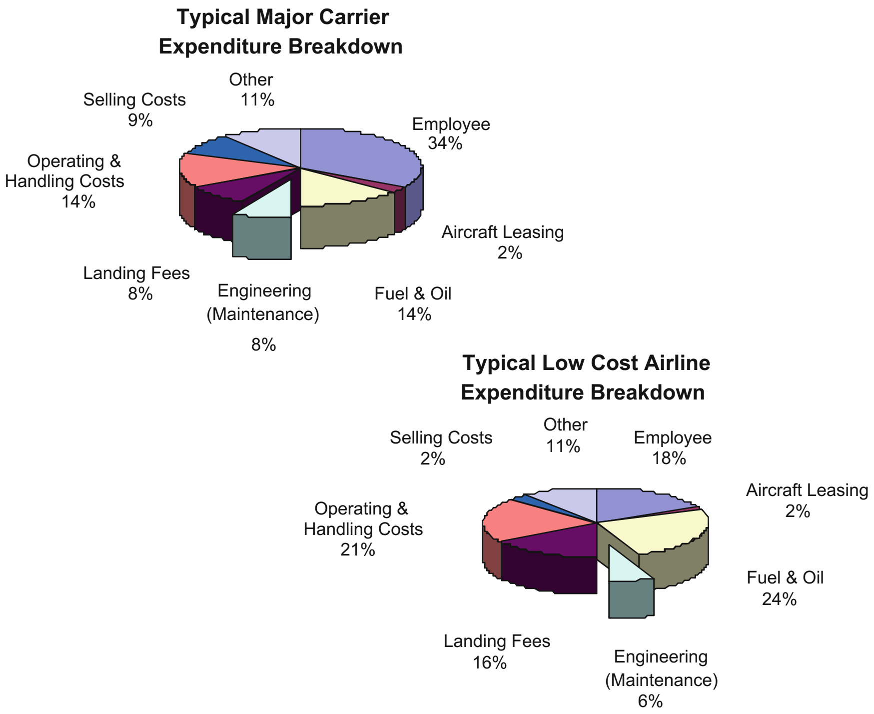


Fig. 5.1 Maintenance costs as a proportion of overall operating cost for typical low cost and major carrier airlines circa 2005

Of these, fuel typically dominates the cost of engine maintenance by as much as 5–8 times in some customer operations [2, 3].[4] Maintenance and support activity clearly has a significant impact on all of the above costs as it impacts the engines health in terms of both functionality and performance efficiency (fuel consumption) (Fig. 5.1).

5.2.1 Indirect Operational Disruption Costs

Disruption costs are incurred as a result of additional work or activity to mitigate the operational impact of loss of functionality of an engine.

Unless covered by warrantee these costs typically fall directly to the aircraft operator.

---

[4] Fuel price increases since these 2003/2006 conference presentations/papers have significantly increased the ratio of fuel price to overall aircraft maintenance cost. Also engine maintenance cost is only a sub-set of the total Aircraft maintenance cost.

For example:

- Lease costs for a replacement engine/aircraft
- Additional Airport landing fees
- Cost of compensation of passengers for delay or cancellation
- Cost of accommodation of passengers for overnight cancellations
- Repositioning of flight crew
- Recovery of aircraft requiring remote location engine change

Such costs are highly variable and dependent upon the location, severity and timing of disruptive events.

Example: Imagine a large civil aircraft carrying 250 passengers that experiences a cockpit warning message during engine startup. The aircraft is scheduled to fly in the last 30 min of the airports operational day before night time noise curfews come into force. The passengers have started the embarkation process and are taking their seats on the aircraft.

- If the message requires a change of an oil or fuel filter (high filter pressure drop detected):
  - A filter change may be possible via a small engine cowl access door within 10 min. The aircraft possibly leaves a few minutes late.
  - If the filter is not accessible via an access door it may be necessary to raise the engine cowls (large doors covering the majority of the side of the engine). Question: as a passenger how would you feel entering an aircraft where the mechanics appear to be 'fixing' the engine? This alone can cause a policy of no 'open cowl' activity within sight of the passengers or during embarkation/debarkation (approx 20 min after landing or before take-off):— The aircraft is in danger of failing to depart before the curfew, incurring overnight accommodation costs for 250 passengers, flight schedule impacts to address out of position aircraft and crew (the stranded aircraft/crew should have been somewhere else in the morning requiring a spare or lease aircraft to fulfill its schedule), missed passenger connections and the re-seating administrative burden. Additionally there are intangible costs of loss of customer confidence and lost future revenue.[5]
- If the message requires a more invasive maintenance activity it is often more cost effective to remove the engine and replace it to allow the aircraft to return to service. Depending upon the location of the stranded aircraft this may involve shipping of a spare engine and work crew to perform the change, leaving the aircraft out of service for 24 h or more.

---

[5] In Japan domestic flights compete with the Bullet Train which enjoys a near 100% on time departure record. Even minor delays (minutes) in aircraft departure schedules have a significant impact in customer perception and loyalty.

5.2.2 Installed Costs

The installed costs are those direct costs incurred on the engine whilst it remains attached to the aircraft. This is dominated by the consumable cost of fuel and oil but does include a level of installed maintenance, both preventative and reactive.

5.2.2.1 Consumables

Fuel is the dominant cost in the consumables arena. Whilst it is not a cost that falls to the engine supplier it is typically a major factor in the customers' engine and aircraft selection.[6]

Fuel cost is dominated by the efficiency characteristics of the basic engine design and the fuel price. Up to 3–5% of additional fuel consumption rate may occur prior to major overhaul as a result of engine deterioration however. This *performance deterioration* fuel burn penalty is an integral function of the rate of performance loss and the frequency and quality of performance recovery at overhaul. Lower loss rates, more frequent overhaul and better recovery at overhaul all benefit through life fuel burn.

5.2.2.2 Installed Maintenance Burden

Installed maintenance is a mix of preventative activity such as routine borescope[7] inspection, filter changes, engine washing (fuel burn recovery) and engine health condition monitoring. The costs are generally related to the frequency, manpower and equipment required but are massively complicated by the mechanics and constraints of operating out of commercial airfields. These costs are often dictated by some fairly simple considerations.

For example consider engine maintenance access. It could be considered good practice to place the fuel filter at the top of the engine where the fuel enters from the aircraft wing immediately before it is split off to feed the fuel injection system as this minimizes the length of pipe run, weight and ultimately fuel burn cost. However next time you are at an airport have a look out of the window and see if you can see any maintenance stands. They are typically a simple selection of working platforms sufficient for one or two mechanics to stand on. They come in

---

[6] For the majority of large civil aircraft the customer airline contracts purchase of the basic aircraft from an airframe manufacturer (Boeing, Airbus and several others) but then has a choice of engine suppliers. Each engine has to have been designed and certified for the Aircraft type, which due to the huge costs involved typically limits the choice to just two engine options.

[7] A boroscope is a rigid or flexible optical inspection device that allows internal inspection of the engine through relatively small access holes strategically placed within the engines external and internal casing structure.

fairly standard heights up to 1.5 ms and are often placed at all the designated maintenance stands around most airports.[8]

What happens if your engine design places its filter 4 meters off the ground? Are you going to demand 2.5 m tall mechanics or a special stand just for your engine? A special stand means either a huge infrastructure cost (stand cost * airport count * average number of maintenance bays) or poor health and safety practices (stands on stands). Either way it is far better to consider the ergonomics and practicalities of the maintenance activity at the design stage.

The frequency and duration of maintenance is the other key consideration. Aircraft have natural windows of maintenance opportunity. There is the 20 min after the passengers offload and the aircraft is being cleaned (before the next set of passengers board), but remember that the passengers don't want to see the sides of the engine exposed and maintenance crews in action.

There is the overnight airport flight curfew when many aircraft are unable to fly for noise reasons. This can provide a 6–8 h maintenance window.

There are scheduled aircraft maintenance windows (like the 10,000 mile or annual service of your car) that can provide even longer maintenance windows.

Quite reasonably aircraft operators want no additional burden or disruption from installed engine maintenance. They want all installed maintenance needs to match the available windows of opportunity, and as infrequently as possible to minimize mechanic and equipment costs. This raises the issue of advanced warning of maintenance needs. The longer the maintenance activity takes the less frequent the opportunity for disruption free maintenance is. Therefore:

- Actions that take 5 min can be accommodated with only a few hours notice (time to schedule the mechanics).
- Actions that take >20 min (overnight window at an airport with suitable maintenance facilities available to the airline) require several flights notice. Note that an individual aircraft may have up to 8 flights scheduled in the day or have a route structure that returns it to its home airport only once a week.
- Actions requiring longer than the overnight window may require several months of continued safe operational time in order to avoid a special maintenance *out of service* event.

Monitoring the engines health to provide such warning is a major requirement to avoid the costs of disruption.

We also have to consider the practicalities of airport life. Some maintenance activity requires the engine to be tested (potentially up to take off power levels) after completion to ensure absolute safety. E.g. after changing some control or fuel system elements. At many airfields this may require the aircraft to be moved to a special location for noise or airport personnel health and safety reasons. It may even

---

[8] The author vividly recalls a personal blinding realization of the complete disconnect between the well lit, CADD enabled design office and the reality of Narita Airports maintenance stands (on a cold wet windy afternoon in failing light) as our airline host politely pointed out the practicality of some of some early conceptual design features.

require the allocation of a normal take off slot on the runway with the loss of airport capacity. This immediately runs into scheduling issues; can the aircraft be moved or does it need a break in other aircraft traffic movements? Can such tests be conducted in the night time noise curfew? How long will it take to reposition the aircraft, test it and return it to the departure gate? These considerations can turn an apparent 30 min maintenance activity into a 24 h elapsed time event.

5.2.2.3 Monitoring Costs

Modern engines are generally equipped with some level of health monitoring capability. At the basic level this may be limited to self checking algorithms embedded within the electronic engine controllers that can display a cockpit message when an error state is detected. At the highest end of sophistication health monitoring becomes a major feat of technology. For example operators of the latest generation Rolls-Royce Trent Engine family which power the most recent Airbus and Boeing aircraft are supported by a 24/7 operations centre based in Derby in the United Kingdom with access to near real-time engine health status.

The health monitoring system is embedded in the architecture of the engine with access to the standard engine control parameter sensor data as well as some dedicated sensors specifically for health detection. Onboard algorithms constantly check these measurements against expected levels in order to detect the earliest indications of abnormal behavior. Detection methodologies vary but tend to centre around absolute deviations from expectation, comparative differences (detector A versus detector B) or rate of change of parameters over time.

Anomaly detection action is then graded according to the type of anomaly:

- Set a maintenance message that will be available at the next maintenance inspection (mechanic plugs in a message downloader).
- Send a message to the pilots display providing them with information to which they may need or choose to react.
- Send a message to a ground station by satellite (during flight) providing information for more investigation by ground based support systems.[9]

Additionally a certain amount of summary data of the parameters recorded during each flight are routinely collected and provided for automatic or requested download when the aircraft reaches an enabled airport.

Once data is available to the ground based systems it is run through a barrage of analysis that establishes a measure of the engines health. Any anomalies that are detected can then either be automatically triaged and the appropriate people

---

[9] Imagine the equivalent whilst driving your car. The onboard monitoring notices an approaching low brake pad thickness indication and it texts a warning message to your local garage. When you arrive at your home that evening the local mechanic is there waiting with a set of replacement pads and all the equipment required to change them for you. You are never exposed to the potential safety issue of poor braking or the additional cost of metal on metal damage to your brake discs.

informed (e.g. e-mail/text message to airline maintenance staff), or can be referred to the 24/7 help desk and their supporting expert support for decisions on the most appropriate reaction to recommend.

Consider the previous example of the filter change being required. Imagine the benefits of in flight or ground analysis based detection of a gradually increasing filter pressure drop. Weeks in advance of the 'do not dispatch aircraft' message the maintenance team is contacted requesting procurement and replacement of the filter. The activity can then be scheduled for a convenient location and overnight maintenance slot with no risk of customer and operational disruption.

There is of course a cost associated with provision of this type of monitoring service in the form of communications costs (satellite and ground based), data warehousing and analysis, continuous development and refinement of algorithms for detection and the staffing of help desk and supporting technical staff. However this generally proves to be low relative to the major benefits in operator schedule adherence and reduction in indirect disruption costs (individual disruption events can be valued in $ millions).

5.2.3 Uninstalled Costs—Maintenance

We can breakdown the uninstalled costs further to the following set. These broadly align with the mechanics of the maintenance process:

- Removal of the engine
- Replacement of the engine (substitute engine asset)
- Transport of the engine
- Disassembly into major sub assemblies (Modules)
- Module disassembly
- Inspection
- Repair or replace
- Rebuild engine
- Test engine
- Transport engine
- Re-install engine

5.2.3.1 Removal of the Engine

Removal costs are broadly associated with the dimensions of time, manpower and equipment.

The time element impacts the costs associated with all the assets occupied during the period of removal. This includes the aircraft itself.

If the engine change (removal and instillation of replacement) can be completed in a 6–8 h shift then it can probably be achieved during a scheduled overnight downtime of the aircraft with no disruption to its planned operational schedule. If it can't then a replacement aircraft will have to be utilized with lease costs (or spare aircraft capacity inventory costs). Alternatively the replacement will be constrained to coincide with other planned aircraft maintenance that already requires the additional downtime.

Manpower and equipment costs relate to the number of events that occur and where they occur.

Take for example the case of a maintenance requirement that causes an aircraft to be stranded away from its home base, at an airport without its own maintenance facilities. A *remote site rescue team* and a spare engine may have to be dispatched to travel to the stranded aircraft, conduct the engine change and then recover both the aircraft and the removed engine back to the home base. The costs then include the travel costs for the maintenance team, spare engine, spare engine stand, removal equipment etc. The travel time alone probably demands the use of a lease aircraft to support the stranded aircrafts schedule commitments.

Removal costs therefore range from a team of 2 mechanics for 8 h through to the costs of leasing an Antonov transport aircraft to ship the spare engine and equipment.

5.2.3.2 Replacement of the Engine (Substitute Asset)

Maintenance of the engine may take as little as a day or two or may extend to months for a major refurbishment overhaul.

During this time the aircraft will either remain out of service or will require a spare engine asset to be installed. Whilst many larger airlines carry their own stock of spare engines to cover this requirement many smaller ones rely on access to lease engine pools. In either case there is a substantial cost of purchasing, storing and maintaining the spare engine pools that is borne ultimately by the airline.

Positioning (global placement) and availability of these spare assets clearly has a major impact on the potential for aircraft to remain out of service longer than is absolutely necessary for the engine change. Failure to have a spare engine available often, inevitably leads to the need for a spare aircraft at a much higher lease cost or cancelled revenue generating operations.

5.2.3.3 Transport of the Engine

The fuel efficiency of a modern gas turbine aircraft engine is related to its bypass ratio.[10] This has the tendency to push engines to ever greater diameters in order to generate a required level of thrust at the lowest fuel burn (and environmental) cost.

---

[10] In a bypass engine only a proportion of the total airflow pulled into the engine passes through the engine core. This is mixed with fuel and is ultimately expelled as a high temperature, high speed

For the highest thrust applications this represents a challenge in transportation. There are three basic modes of transportation for an engine. One is in place on the wing of the aircraft, second as freight on an aircraft and third as ground freight.

The first of these is rarely utilized as it generally precludes the aircrafts use during the flight for commercial purposes and it presupposes that a replacement engine already exists at the destination.

The second is subject to the availability of transport aircraft. Military transporters tend to have the highest payload capacity (size and weight) but are not generally available for commercial airline hire. Some specialized commercial cargo aircraft are available for large loads but these are expensive to hire and have limited availability. Therefore the desire is to be able to transport the engine in a widely available commercial cargo aircraft such as the Boeing 747. This has a prescribed limitation on the size of the freight due to the dimensions of the loading door and internal airframe capacity. Failure to design within this size window dramatically increases transportation costs. Some designs of engine require the bypass fan assembly to be removed from the rest of the engine in order to be able to meet the transportation size constraints (although there are also other considerations of modular overhaul frequency requirements that impact on this design decision).

The third option of ground based transport has its own issues. The first and foremost is clearly one of speed. There are however also practical constraints. For example road freight out of some major cities requires the engine to be transported through road tunnels (e.g. New York) or similar with dimension constraints. These result in time window limitations (you can only transport loads above these dimensions overnight) or size limitations (road freight in Tokyo restricted to 3 m square section). Both can impact design considerations.[11]

5.2.3.4 Modular Disassembly

The engine has now made it to the overhaul facility and undergoes the first stage of disassembly.

The important thing to recognize is that most complex mechanical designs have a degree of modularity where the product can be progressively broken down into smaller subassemblies until it eventually reaches the level of individual components. This is true of modern gas turbine engine design.[12]

(Footnote 10 continued)

exhaust (the traditional Jet engine). The jet core flow in a bypass engine is kept to a minimum and is primarily used to generate the power necessary to turn a much larger set of fan blades that push a slower but much higher volume of *bypass* air from the rear of the engine. The higher the bypass ratio the more efficient the engine becomes in simplistic terms.

[11] The Trent 1000 engine for the Boeing 787 Dreamliner aircraft has a special flat on its fan case flange that allows it to meet the 3 m transportation requirement for Japan.

[12] See the Rolls-Royce publication [4] or the web sites of most of the major Gas Turbine Engine manufacturers for examples of engine modularity.

In this first phase of disassembly the engine is stripped down to a level that allows the modules requiring maintenance to be accessed. Modules that do not require maintenance (at this point in time) may be stored awaiting the refurbished modules to be completed or may be re-allocated to another engines rebuild.

Importantly at this stage the modularity dictates how much of the engine will be subject to the next level of disassembly activity. Take the hypothetical example of an engine with just two modules. The first is the main jet gas path which sees the highest temperatures, pressures, speeds and loads within the engine. The other is the bypass system which lives in a relatively benign environment (speeds, temperatures, pressures). You would rightly expect that the gas path components would deteriorate at a higher rate than the bypass and would therefore require more frequent refurbishment maintenance. If all the high deterioration parts are contained in the jet gas path module then you would expect to overhaul it more frequently than the bypass.

e.g. a maintenance policy such as.

Overhaul jet gas path → overhaul both → overhaul jet gas path → overhaul both.

But what happens if you include just one high deterioration part in the bypass module as a result of poor selection of the module interface line? Either the part does not have the life expectancy to reach every second overhaul event as planned (disruption from failures, reduction in function or unplanned premature maintenance) or the module has to be planned for overhaul at every occasion (additional strip, inspection and build costs for all the components that did not required overhaul at this interval.

The decision on which modules to expose for module disassembly is one based upon both the timing and reason for the overhaul event as well as the basic modular architecture and maintenance policy for the product.

As already discussed when an engine is put into service there will be an expectation of a *reasonable* operational life before major overhaul maintenance is required. If an unexpected maintenance requirement arises well before this life a decision is required on the purpose of the maintenance that is done. Take the example of an engine that has achieved just 40% of the expected time to overhaul before one of its sub-systems develops the early symptoms of a fault. There are two choices:

1. Do sufficient maintenance to address the immediate need with the intention of allowing the engine to complete the other 60% of the intended life to its refurbishment overhaul (*check and repair*).
2. Pull forward the planned refurbishment and undertake sufficient work to return the engine to service with a full 100% life before the next refurbishment expectation (*refurbishment*).

This is typically an economic question based upon the cost of the *check and repair* option versus a full *refurbishment* overhaul. If the *check and repair* costs 10% of the full *refurbishment* then it might be reasonable to undertake this approach at a life up to 90% of the planned overhaul interval but not beyond. In practice this type of logic is normally embodied in a set of module overhaul policies which often employ a concept of *soft lives*. The *soft life* is the life beyond which the module would normally be refurbished if a maintenance opportunity is presented.

At induction to the overhaul facility a typical modular overhaul decision making process tends to be:

- Is the module beyond the policy soft life? If yes then refurbish.
- If below the soft life conduct basic modular inspection to ensure its health is in line with expectation for its life. If inspections failed then refurbish.

5.2.3.5 Module Disassembly

Modular disassembly can be simple (strip to individual components), or complex (strip to sub assemblies and then apply logic as for module strip decision). In either case it may involve a large amount of labour and time to reduce the modules down to the level of piece parts or subassemblies that can be inspected and refurbished. The design of the product can have a profound effect on the costs and complexity of disassembly. For example interference fit spigot joints[13] may be easy to make on assembly but provide a host of difficulties on disassembly that result in additional damage or even component scrap risk.

5.2.3.6 Inspection

There are three basic outcomes of the refurbishment decision making process at piece part level.

1. The part is acceptable to be refitted without any other work being done. i.e. the part has sufficient residual life to allow the next panned overhaul period to take place without significant risk of the part failing to perform its function.
2. The part does not have sufficient residual life to meet the criteria for 1) but can be economically repaired and would then meet the residual life criteria.
3. The part has insufficient residual life and is beyond economic or effective repair limits.

---

[13] Some joints need to be extremely tight to provide mechanical stability of the assembly. One such joint type involves two concentric rings on mating components in which the inner rings outer diameter is larger than the outer rings inner diameter. In order to assemble the two parts the inner ring is shrunk by super cooling, often with liquid nitrogen, in order to temporarily remove the interference. As the inner part warms, it grows back into contact with the outer part. Once the two are in close thermal contact the reverse operation becomes much more difficult. It may prove impossible to shrink the inner component leaving mechanical force as the only method of disassembly.

The purpose of inspection is therefore to make this determination in the most cost effective manner. This is often accomplished by working through a series of inspections that determine the acceptability of the functional features of the part. Careful consideration of the order of such inspections can optimize the costs of rejecting a part. E.g. if 90% of all rejections take place at inspection number 10 then can the inspection be done as step 1 saving the costs of the previous 9 steps?

The counter argument to the optimized inspection order is the one of providing information to enable product improvement. If 90% of parts are rejected at inspection step 1) is it cost effective to develop a repair for this feature? If only 2% of parts would be rejected for the other feature inspections then the answer is possibly yes. If however 80% would also be rejected for subsequent inspections then the single repair would yield few benefits.

Other issues that can occur are the result of the differences of inspecting for refurbishment versus new product assembly. Simple issues such as part marking and measurement datum arise. If either is placed on a surface or feature that is subject to wear then you may find that a perfectly functional part cannot be refitted because it can no longer be identified with 100% confidence or the reference point for a critical dimension has been compromised. Often all these issues require is some consideration of the build, strip, inspect; build, strip, inspect cycle at the point of design.

5.2.3.7 Repair or Replace

So we now have a part or assembly that has been rejected for refitting. The decision is do we repair it or replace it?

If we replace it we invest the full cost of a replacement part and the full environmental cost of disposal or recycling of the rejected part. However we will be obtaining a part that has the fullest possible residual life (as it is new).

If we repair it then in all likelihood we are only repairing certain features. Whilst these may be returned to the as new life, those features that we don’t repair will not. Dependent upon the combination of features and their expected residual lives, a repaired part may have considerably shorter residual life expectancy than a new part. Additionally, repair often involves additional processing relative to the new production part. For example repair of coatings often requires the original coating to be removed before a new coating can be applied. This involves additional costs both financially and environmentally.

The repair or replace decision is therefore a complex one that is ultimately best made based on the detail of the individual component or assembly. However, at the point of design it is essential to have a view of the likelihood of repair being the viable economic and environmental option. If it is, then there are design decisions that can be taken that maximize the probability of a successful repair whilst reducing the cost and environmental impacts.

5.2.3.8 Rebuild Engine

Rebuilding the engine after refurbishment should be a repeat of the new product assembly process. However because the engine may not have been fully stripped to individual piece parts there may be some differences.

5.2.3.9 Test Engine

Dependent upon the nature of the maintenance work undertaken a *pass-off* test is typically required to demonstrate that the engine is fully functional and meets guaranteed performance levels prior to return to the customer. Such testing is generally conducted at a dedicated test cell at or near the overhaul facility.

5.2.3.10 Transport Engine

A repeat of the previous transportation considerations.

5.2.3.11 Re-install Engine

The reverse of the previous removal process.

5.3 Is a Gas Turbine Engine Unique in Its Service Support Complexity?

As can be seen from the previous sections the costs associated with operating a gas turbine engine for large aircraft is complex, with a multitude of degrees of freedom that impact those costs.

However none of this is particularly novel to the gas turbine example. The majority of considerations apply equally to a range of product sectors. For example take a washing machine. Other than the presence of rotating and static elements in the mechanics it apparently has little similarity to a gas turbine. However many of the considerations for service cost are the same:

- Do I react to the strange noise the drum is making or wait to see if it stops working? (*health monitoring*)
- When it fails will it just stop or will it ruin my clothes/flood the floor? (*indirect disruption costs*)

- Will I call out the repair man or is it old enough to make it cost effective to invest in a new washer? (*check and repair* or *replace*)
- If I go for repair will they replace the timer module or just the obviously burnt out capacitor? (*modular* versus *component* overhaul).

5.4 The Nature of Cost

In order to be able to deal with the complexity of the cost, it is helpful to be able to construct a simplified mental model.

$$\textit{Total cost} = \textit{intrinsic cost} + \textit{inefficiency cost}$$

5.4.1 Intrinsic Cost (Minimum Cost with Perfect Decision Making)

- All mechanical systems exhibit wear out as a result of use
- With perfect visibility and understanding of this wear out there is a minimum fundamental cost of maintaining a functional product.
    - Maintenance undertaken at exactly the optimum time
    - Precisely the right quantity and location of maintenance resources to meet the demand
    - Just the right level of maintenance done to achieve optimum functionality/cost of support.

5.4.2 Inefficiency Cost (Additional Costs from Imperfect Decision Making)

- Visibility and understanding of wear out or other causes of damage is imperfect
- Decision making is constrained by factors such as understanding, training, time available to decide etc.
    - Maintenance undertaken too late—disruption or secondary damage incurred
    - Maintenance undertaken too early—useable operational life is lost for the same maintenance costs
    - Insufficient maintenance resources—maintenance delayed (as above) or queued requiring additional spare assets

  - Excessive maintenance resources—additional inventory and employment costs
  - Insufficient maintenance done when opportunities arise—short life to next maintenance event equals loss of useable operational life
  - Excessive maintenance done when opportunity arises—additional costs for no additional operational life benefit

Intrinsic cost and inefficiency costs are both fundamental. Unless the product is never user it is impossible to reduce either to zero. The trick is to do what we can to minimize both elements. Consideration of both allows us to simplify the discussion of how to minimize service maintenance costs whilst considering both the product design and service delivery elements:

1. MINIMISE INTRINSIC COST: optimize the design to maximise the operational life and minimise maintenance requirements.
2. PREDICTABILITY: optimize the understanding of maintenance requirements to maximize the probability of undertaking the right maintenance at the right time with the right amount of available resources.
3. MONITOR: measure and record actual behaviour to refine understanding and improve the quality of decision making.
4. REACT: address opportunities to improve either the intrinsic cost or the inefficiency cost as appropriate.

Methodologies for addressing these are discussed in subsequent sections.

5.5 The Nature of Deterioration

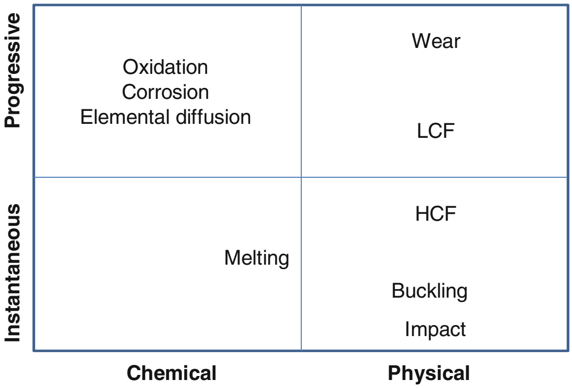
Before discussing opportunities to improve intrinsic and efficiency costs it is worth a quick examination of the root cause of all of them all; deterioration.

The only reason for undertaking maintenance is that products deteriorate with use (or over time). If no deterioration took place the product would function for ever exactly as it did when first manufactured and the maintenance industry would all be out of a job. However there are no perpetual motion machines and neither is there a deterioration free product (even the most passive products can be subject to external sources of damage).

So let us examine the nature of deterioration. There are a number of facets to the process of deterioration all of which impact our methods of tackling the subsequent maintenance costs.

- Progressive versus instantaneous (e.g. gradual wear versus impact damage from a bird flying into the engine)
- Short versus long time-base (e.g. does significant deterioration and loss of function potentially occur within a fraction of a standard overhaul interval or over the period of a number of overhauls. In the latter case there will be one or

Fig. 5.2 Illustration of just one classification method for deterioration mechanisms



	Chemical	**Physical**
Progressive	Oxidation Corrosion Elemental diffusion	Wear LCF
Instantaneous	Melting	HCF Buckling Impact

more opportunities to observe the progression of deterioration and take remedial action)

- Monitorable versus hidden (e.g. slow measurable change in gas temperatures over months versus sudden failure of a part with no prior warning)
- Utilization sensitive versus utilization insensitive (e.g. thermal degradation of parts closely related to the number of take offs of the aircraft versus corrosion of external parts which may progress even if the engine remains unused)
- Fundamental versus probabilistic (e.g. any two mating parts will wear to some level [100% probability] but significant wear may require some other factor that is not always present [less than 100% probability]) (Fig. 5.2)

A further key assertion is that in the vast majority of cases deterioration impacts features of components or systems rather than the whole physical entity. Whilst we may talk about failure of component X or repair of component Y what we actually mean is that a feature of component X has deteriorated to a point where the component no longer serves its function or a feature of component Y has deteriorated to the point where it needs repair (but other features of the part may remain perfectly functional).

The reality of any physical system is that it is a collection of systems, subsystems, parts and part features that together combine to perform some overall function. Deterioration principally occurs at the feature level. Each feature may suffer from many or no significant deterioration mechanisms, each of which competes over time to undermine or reduce the functionality of the feature. When sufficient loss of feature functionality occurs (either individually or in combination) then the overall system function is impacted requiring either proactive (avoiding) or reactive (recovering) maintenance.

Managing maintenance costs (direct and indirect) is all about learning to predict, avoid, mitigate and recover from feature based deterioration.

5.6 The Design and Verification Process—Mitigating Deterioration

So how do we tackle the problem of optimizing maintenance cost when we are faced with such complexity of causes and potential methods of treatment?

The answer is to reduce the problem to its fundamental elements as far as possible.

1. Do we have a clear understanding of the requirement? Do we want the minimum possible direct maintenance costs (overhaul bills), or do we want the minimum overall cost of operation (fuel costs, installed maintenance and overhaul cost)?
2. Do we have a clear strategy for achieving this at the whole system level? Do we have clear product capability requirements (time between overhaul, operational reliability levels, cost of typical overhaul) and a corresponding support policy and infrastructure (fix when non-functional vs fix before loss of function, overhaul whilst in situ or return to overhaul facility)?
3. Do we understand the key success drivers of this strategy (physical deterioration mechanisms, optimum maintenance policies etc.)?
4. Can we mitigate the risks of loss of function at the optimal point in the process (avoid, detect, contain, recover)?
5. Can we predict expected behaviour of the product/service and monitor reality against it (early identification of issues)?

This amounts to the classical V of requirements cascade, design action and solution verification (Fig. 5.3).

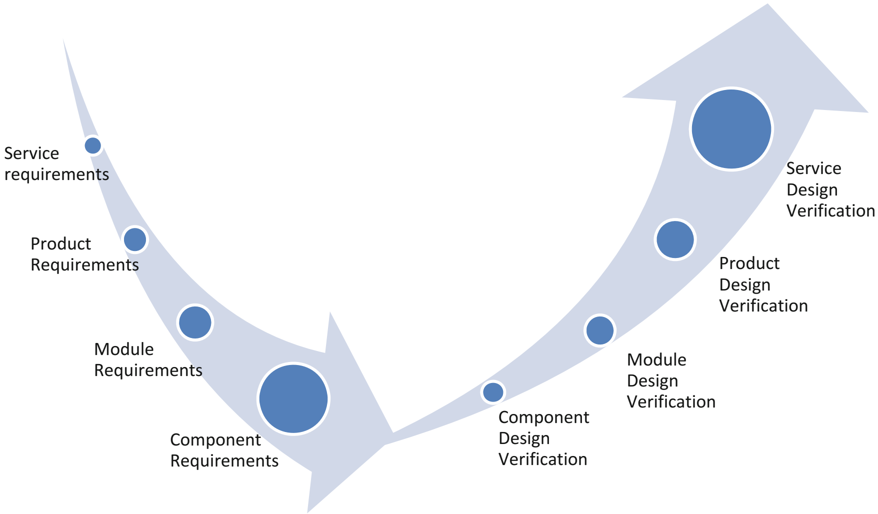


Fig. 5.3 The V model of requirements breakdown, design solution and verification

We will consider each of these stages in turn to examine the types of action that can be taken to optimize product design and maintenance service offerings.

5.6.1 Requirements Cascade

This is arguably the most critical of all the stages in terms of determining overall success. It is worth at this point considering the plight of the design engineer charged with the definition of one component within a complex system such as a gas turbine engine.

The design engineer will be faced with a series of requirements that they are expected to meet. These will come from a variety of sources; some may be legal requirements to meet safety regulations, some may be derived from customer requirements and others will be internal business requirements from their own company. In all cases the likelihood that the requirement will be met has three primary dimensions:

- Do I understand what I am being asked to deliver?
- How important is it relative to other requirements?
- How easy is it to achieve?

Maintenance cost traditionally suffers on all three counts:

- How do you express the maintenance cost requirement at the level of the component? The costs are typically billed at the whole system level, or if broken down typically refer only to the direct costs incurred in repair or replacement of the part. How do you define the requirement for a component that costs just $5 to replace, but causes premature overhaul of the entire system at a cost of $1 m? How do you remove the ambiguity of the frequency of overhaul of the part the designer is accountable for being determined predominantly by the lives of other parts of the system?

Example A component has a value of $5 and an engine contains a set of 100 parts. All parts are replaced at routine overhaul at a cost of $5 * 100 parts + $50 labour cost. The typical cost of refurbishing the rest of the engine is $1 m per overhaul. The frequency of overhaul is entirely determined by the part in question (it consistently has the shortest life of all the parts in the system). If the part averages a life of 10,000 h before overhaul is the costs per operational hour due to the part:

(a) $550/10,000 h = $0.055/hr (direct cost expended on the part)
(b) ($550 + $1 m)/10,000 = $100.055/hr (direct costs *caused* by the part)

What would the cost be if the engine had a second component that would have caused overhaul at 11,000 h?

What would the cost be if each of the 100 parts in the engine had a distribution of life around 10,000 h such that on average the *set* required overhaul more frequently than the 10,000 individual part average life?

What would the cost be if the mode of failure caused operational disruption with additional costs to the customers operation over and above the pure maintenance recovery cost?

- Safety considerations are absolute requirements for product certification. Weight is a hard customer requirement (as the wing and pylon are certified to a maximum engine weight) and unit cost is a key requirement for business viability. All have feedback loops within the timescale of the design process. Any failure to achieve requirements then becomes apparent and subject to escalation of priorities. Even when requirements start with equal priority, those that have an effective feedback loop of status versus requirement quickly get brought into much sharper focus and generally get higher levels of attention.

Example Component x needs to achieve (a) a weight of 1.5 lb, (b) be purchased for less than $650 and (c) have a maintenance cost of less than $0.09 per hour of service operation.

Weight is predicted with high confidence from the CADD drawing package in real time.

Supplier quotations are available within weeks

Maintenance costs may become apparent 5–10 years into service (but unlikely to be broken down to the level of individual components).

- The design changes necessary to impact maintenance cost are in reality no more or less difficult to accomplish than for the other requirements. The difficulty appears in the ability to understand the relationship between the options available to the designer and the impact they are likely to have on the maintenance cost. The major source of this complexity is the statistical nature of deterioration and the many uncertainties in the driver; e.g. *if* the operator uses it as we expect; *if* it is maintained as we expect etc.

Example Selecting a more expensive and denser material with superior properties *may* increase the life for a particular deterioration mode but is this the only deterioration mode that is significant in determining the parts ultimate life? The change in material will *definitely* impact the component cost and weight.

What the design engineer therefore needs is a set of clear and unambiguous requirements. The role of the requirement is to frame the problem in such a way that it reduces the issues above to the minimum whilst ensuring that all the designers and other stakeholders are aligned to delivering a total system solution that works. It is no use having a system of 20,000 parts where 19,999 meet the 10,000 h life requirement and the remaining one is only capable of 8,000 h. Neither is it acceptable to design a product for 10,000 h maintenance interval with a maintenance network sized for a 12,000 h interval.

5.6.2 Prediction and Design Action

The absolute key here is our ability to generate sufficient understanding of the impact of our design options on the maintenance cost to be able to guide effective decision making. If we are uncertain of the outcome of choices we will, as a direct result of human nature tend to prevaricate, postpone or otherwise avoid making the decision. Gambling ability is rarely part of the designer recruitment process!

The single biggest thing we can do to improve the consideration of maintenance costs in the product design is to make it feel real to the design population. It is the authors experience that the design community likes nothing better than the opportunity to solve a well stated and justified problem. Provided you give them the clear problem statement, sufficient understanding of the issues and the time, freedom and tools to explore the problem they will come up with truly great solutions.

So what are the magic ingredients?

- Education: make the problem *feel* real and important.
- Education: explain the complexity but bring it back to simple achievable goals that drive >80% of the costs.
- Make the requirement numerate or a clear pass/fail criterion.
- Predict the maintenance cost—waiting for reality is doomed to failure, having a mechanism to raise priority when failure to achieve requirement is predicted is key.
- Treat the prediction as reality; accept that it is less certain than other attribute predictions and measurements but don’t use that as a rationale to ignore it when it looks hard to address.

Have you noticed the trend in the items above? With the exception of the maintenance prediction capability, the majority of issues are actually related to influencing human behaviour. If you want to be effective in addressing intangible hard to define, statistically difficult requirements with exceptionally long lead times to the emergence of proof of success or failure then it is far more a behavioural problem than a process or tools issue.

> You can’t design for maintenance cost successfully unless you address the human behaviour issues. Tools and processes alone are doomed to fail.

Some of the practical techniques available are therefore:

- Expose the design teams to the realities of service life. Ideally directly though attachment, secondment or visits but at the very least through direct contact with the front line staff.
- Recognise that the customer often has the most experience, so find a way to access it.
- Work through historical issues and understand the root causes and drivers of service impact. Take this learning through into the earliest stages of design consideration.[14]
- Conduct failure mode analysis and design to avoid all the associated consequences including maintenance cost.
- Enumerate and communicate the predicted status against the requirement and react to predicted shortfalls.

5.6.3 Feedback Loops

Service feedback is critical to success in managing maintenance costs. The more complex the product becomes the more fiendish the combination of deterioration mechanisms leading to service costs appears to be. It is practically impossible to predict all possible causes of service deterioration, rate of functional impact and the subsequent cost of management and recovery.

Feedback loops are therefore vital to obtaining the earliest possible identification of an emerging cause of cost and allowing the maximum reaction time to address it.

Multiple feedback loops are possible within the life cycle of the product:

1. Feedback from prior experience (review of similar product histories).[15]

   Note however that differences in environment, design or utilisation may impact the relevance of this experience.

2. Peer or customer review (identification of potential service cost drivers by those with more direct service experience and knowledge than the design team themselves).
3. Simulation and prediction (generation of a predicted cost using best available predictive methods).

---

[14] The majority of organisations will conduct a FMECA or Failure Mode Effect and Criticality Analysis as part of their design/safety process. In your organisation is this done before design starts as a guide to good design practice or after design as a check for safety compliance? One of these approaches has a high probability of reducing maintenance costs relative to the other.

[15] With the time from product design launch to statistically significant service maintenance data being measured in decades it is often the case that the only solid data comes from a design of engine 2 or 3 generations older than the current project.

4. Component or system test. This can be to demonstrate functionality at a specific set of conditions or often of more value is the test to failure that demonstrates the limits of capability and level of design margin.
5. Service health monitoring to detect the initiation or occurrence of loss of service function.
6. Service failure and maintenance cost statistics.

Each of these has a valid role in the provision of feedback to the design process. In general the later in the list an item appears the more definitive or real the feedback it provides. The counter point is that it also occurs later and therefore provides much less opportunity to minimise the costs. Items 1–3 generally allow you to impact the intrinsic cost of maintenance as they allow the opportunity to change the basic product design. Item 4 may offer the same opportunity but quite often test opportunities occur late in the design process when the ability to introduce change is constrained. Hence items 4–7 predominantly allow you to impact the inefficiency cost through provision of improved understanding of intrinsic deterioration costs and therefore the opportunity to manage them more effectively.

Also consider the customer impact. The final two items are the point at which it's the customer who feels the pain of unexpected service deterioration. Finding and resolving a potential issue from an earlier feedback loop can eliminate it before it ever gets seen by the customer[16] avoiding customer dissatisfaction and the inevitable impact that has on future revenue opportunities.

5.7 Whose Costs Should We Design for, Ours or Our Customers?

I am often asked which of the service costs should we actually design for? After all it's the customer who will pay for fuel and indirect disruption costs but we will pay for the costs of maintaining to avoid them (assuming a fixed price maintenance contract has been agreed).

There are two ways to look at this, both of which lead to the same basic conclusion:

- Unless you intend to go out of business you must be hoping that your current customers will in the future continue to buy your products and services. In their shoes would you prefer to buy from a supplier who actively helps you reduce all

---

[16] Whilst customers can be very vocal about the positive benefits that health monitoring brings it is only ever minimising the pain that is inherently built into the product design. Most customers would much rather have you design that pain out before it ever gets to them. Rolls-Royces major customers all agree that whilst rapid and effective response to their emerging issues is an admirable trait in a supplier they would rather operate boring and predictable products that never present issues in the first place.

your costs or one who makes decisions in their own interest even if it's at your expense?
- Most businesses exist to make money. The opportunity is represented by the difference between the revenue available (which is a measure of the value it brings to the ultimate customer) and the costs of providing the product or service. In the civil aviation market that means the revenue depends on the ability to get passengers and cargo reliably and comfortably from point A to point B. The costs include acquisition of aircraft and engines, fuel, crew, disruption recovery etc. The bigger the gap between the revenue generating value and the true costs the more profit is available to share between all the parties involved in the value chain.

It may be slightly heretical but my personal view is that the engineering task is to create the largest possible profit pot; the commercial teams' task is to agree how it gets shared between the organisations involved.

So the answer I inevitably come to is that it's in the organisations long term interest to maximise the value the product and service brings whilst minimising the total cost of their provision. If this is apparently at odds with the organisations immediate best interests then set the commercial team the challenge of capturing a fair proportion of the profit margin that is being generated.

5.8 Conclusions

In this case study of a gas turbine engine what have we concluded about the process of designing for service maintenance costs?

Firstly the very nature of service adds many dimensions of uncertainty to the problem, such as how will the product be used? How will it be maintained? How will two seemingly identical products deteriorate when exposed to the same environment and operation? This complexity and uncertainty is at the root of why maintenance costs traditionally get less attention in the design process that other more tangible product attributes.

Secondly, the problem is essentially one of human behaviour. It is the difficulty we as human beings face in dealing with this complexity that inhibits our ability and willingness to tackle the problems effectively.

We have however discovered that there are ways to almost artificially simplify the situation and reduce the intangibility that inhibits effective design. Clear objective requirements, effective education (and simplification) of the service realities, creation of successive feedback loops all allow our design communities to engage more effectively with the issues. It is not an easy transition to make from a product design company that also sells support services to a service company that happens to design the underlying product, but it is rewarding.

The final conclusion is that by looking beyond our own revenues and costs to those of our customer we may discover opportunities to generate additional value generation opportunities.

Acknowledgements Some text from this chapter is reproduced with the permission of Rolls-Royce plc from a future planned issue of *The Jet Engine* book.

References

1. Rolls-Royce plc Investor news (2009) Paris Air Show orders reflect strength of Rolls-Royce portfolio In: Rolls-Royce corporate website, http://www.rolls-royce.com
2. Horder P, Senior Vice President, SH&E Ltd (2003) Managing aircraft maintenance costs conference (Brussels)
3. Harrison A (2006) GT2006-90570 Design for service—harmonizing product design with a services strategy. In: Proceedings of GT2006 ASME turbo expo 2006: power for land, Sea and Air, May 8-11, 2006, Barcelona, Spain
4. Rolls Royce plc, The Jet Engine book, The Jet Engine—a complete overview of the modern gas turbine. Key Publishing Ltd

Chapter 6
The Knowledge Management Perspective

Charles Dibsdale

> The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge.
>
> —*Stephen Hawking*.

Abstract An important element of product lifecycle management (PLM) is knowledge management (KM). KM helps manage risks inherent in products as they increase in complexity, and the organisations and teams who design build operate and support the products may be dispersed in geography and time. Economic pressures are also forcing organisations to do more for less with fewer resources in reduced time. It is essential that knowledge is exploited if these efficiencies are to be made. This chapter explores problems with Knowledge Management, posits definitions that may be useful about the nature of knowledge and its relationship with data and information.

6.1 Problems with Knowledge Management

Knowledge management is hackneyed and may be regarded by many as a management fashion that is declining Grant [6]. Many who label themselves as KM system vendors are from the Information and Communications Technology (ICT) domain, where their systems enhance electronic communications, data processing storage, linking and retrieval. Although these ICT systems are important and potentially valuable they only address a fraction of what needs to be considered when implementing or running a KM system.

Many people believe knowledge can be treated as an object that can be externalised from people. Objects can be managed, and therefore the whole term knowledge management is a valid concept if objectification of knowledge is possible.

One of the foundations of knowledge is the relationship between data, information and knowledge. These concepts are ill defined and the terminology used in them is often interchanged. Many people also hold that ICT systems are able to produce information and knowledge. If ICT KM systems are dissected, it is plain

---

C. Dibsdale (✉)
Ox Mountain Ltd, Torfaen, UK
e-mail: charlie.dibsdale@gmail.com

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_6

that the ability of machines is limited to data processing, inferring new data or control (via a set of encoded rules). This is true when the latest state of machine learning or artificial intelligence is considered.

An important aspect of KM is that Intellectual Capital and experts need to be identified within an organisation. This ensures an organisation can ensure the preservation of knowledge in its experts. However, by its nature tacit skills may be difficult to recognise.

We will explore the true nature of knowledge and show that knowledge is intrinsically associated with people, and that 'knowledge facilitation' may be a better concept if organisations want to exploit knowledge for competitive advantage. If you are of the opinion that a KM system implementation is purely a technology problem that can be safely delegated to the IT department, you need to read this chapter.

6.2 What Is Knowledge?

In order to understand how KM should be applied, it is necessary to describe the nature of knowledge itself, and review some existing models that help define and provide a solid basis for development.

Knowledge is often described in two forms, called tacit and explicit knowledge.

Explicit knowledge is that which may be encoded and written down, it may be expressed, consciously rationalised and treated as an object. This form of knowledge can be externalised from the human mind, stored, shared and re-used by different people. Tacit knowledge is skilful action based on experience and knowing how to do something. This knowledge cannot be encoded or written down, but facets of skilful actions may be expressed and discussed in the company of a set of skilful practitioners who often develop their own set of semantics that describe the nuances of their tacit knowledge, but this set of semantics does not externalise tacit knowledge, and tacit skills cannot be learned from it in isolation.

Polanyi [11] is attributed for initially defining tacit knowledge uses an example of riding a bicycle as a tacit skill. The tacit knowledge of how to ride a bike is learned by practice, and not by reading a manual. Tacit knowledge is internalised inside humans and cannot be externalised and treated as a separate object. Experts who coach beginners may facilitate tacit knowledge sharing. For millennia, masters and apprentices have practiced this form of sharing tacit knowledge and learning. The medical profession also teaches new doctors by extensive supervised practice in hospital wards before they qualify. Doctors are also very open about consulting with each other, with no loss of reputation.

Tacit and explicit knowledge should not be considered as separate entities, Tsoukas [12], as all knowledge comprises both elements in differing degrees. It is generally accepted that all knowledge must initially be tacitly based (for example reading an academic paper requires the tacit skills of reading, practice in the scientific method and comprehension before the codified explicit knowledge may be appreciated and the experiments repeated).

If knowledge is comprised of Tacit and explicit elements and tacit knowledge cannot be codified (or treated as if it is an object) then how can knowledge be managed? A more useful term would be to use the term ‘knowledge facilitation’.

Explicit knowledge that is codified and recorded should be regarded as data. It is a subset of general data that has been consciously selected and structured with a context in mind by the author. It may be selectively reviewed because of the contextual links, but as a codified corpus, it is not information or knowledge because it is separated from the sense making ability that is unique to human beings. In order to illustrate this the relationship between data, information and knowledge these concepts should be defined. If we research literature for existing definitions, we may be dissatisfied with what we discover.

Data information and knowledge are often interchangeably used or are mixed within a number of definitions. An example of where knowledge management is expressed in terms of, and sounds like, information and data management is taken from the ITIL standard, where KM is recognised as integral to service support of ICT systems:

> The purpose of Knowledge Management is to ensure that the right information is delivered to the appropriate place or competent person at the right time to enable informed decisions.
>
> The goal of Knowledge Management is to enable organizations to improve the quality of management decision making by ensuring that reliable and secure information and data is available throughout the service lifecycle.

This passage assumes information can be encoded and delivered to ‘a place’, and mixes information and data with little distinction. The reference to a competent person implicitly recognises that they possess the necessary tacit knowledge to act.

The Oxford English Dictionary defines Information as:

> the imparting of knowledge generally.

These definitions contain circular arguments, are inconsistent and unsatisfactory.

This confusion has been deepened by marketing where computerised KM systems are attempted to be differentiated from others by claims they can manage knowledge. This is hype, tacit knowledge cannot be treated as an object and it is therefore doubtful whether it can be managed. Knowledge facilitation may be a better term to use in trying to exploit knowledge.

In reality computer systems may only process data, enabling rapid storage processing and retrieval of data. Computer systems may also be controlled by sets of rules, statistical techniques and the application of machine learning to produce other highly contextualised output data that may be used for automated control or enhanced decision support, but this does not make it a system to manage knowledge or information. An information or knowledge system involves people.

The essence of information is that a human being makes sense of data in a context. A human may synthesize different information from the same data in different contexts. Different humans may also synthesize different information from the same data in the same context. We all interpret data differently. Human sense making depends on several factors, namely, the person's experience and tacit knowledge, their values beliefs and morals, their context and their situation. Any changes in these factors may result in different information being synthesised from the same data.

In a business domain, when people generally follow well used defined processes with commonly understood goals, then relevant data within that context may be interpreted by different staff familiar with the domain with minimal variation. This is why ICT based KM systems can be valuable.

6.3 The Data Information Knowledge Wisdom (DIKW) Model

In order to illustrate the relationship between data information and knowledge more clearly the Data-Information-Knowledge-Wisdom model first described by Ackoff [1] will be analysed and critiqued. An alternative amended model will be proposed that better describes the relationships and how the role of ICT may better appreciated.

The following observations are offered to help define important terms and concepts

- Data are unrelated facts, however data is not restricted to text or database entries, it may also include anything we perceive through our five human senses.
- Information is data sensed by human beings, where meaning is derived in a context to gain insights. In this description a wider context should be taken for what data is, how it is sensed and interpreted. A smell (data being sensed) may be interpreted as an indication that food is being prepared, where the information in a known context may be that our next meal is nearly ready for consumption and it is time to break off from our current activities to dine. Knowledge that uses this and other information may be that from previous experience once the earliest smells of food have been sensed one should proceed quickly to the food dispensary in order to avoid queuing.
- Knowledge is the repeated use of information to achieve outcomes through thought or purposeful actions; expertise (or tacit knowledge) may be developed by repeated practice to achieve outcomes more effectively or efficiently. Our minds use associations to store our experience of the world.
- Wisdom is the skilful application of knowledge for optimising benefits, with some authors attributing morals, values or religious connotations to wisdom. Some Posit that wisdom has connections to having a soul and use this as a basis to argue that only humans may have wisdom.

The persecution of Galileo and Copernicus provides enough evidence for the author to summarily dismiss the quasi-religious basis of wisdom.

There is also a further problem with wisdom as being a separate entity from knowledge. It may be regarded as a value judgement on knowledge instead of a true transform of knowledge. Wisdom may also be highly contextual and from different perspectives, one person's wisdom may be judged as another's folly with both parties having justifiable reasoning from their own perspectives.

The transform between data and information is also problematic when information and communication technology (ICT) is considered. It is widely believed that ICT is able to process and manipulate information, where it may be more accurate to say ICT is able to process subsets of data that are deliberately selected to be relevant in defined contexts. The selection and contexts may be described explicitly as codified rules that computers obey. Highly relevant subsets of data to a context are not information, as ICT machinery do not have the ability to make sense of the data. The sense making is the essential ingredient that transforms data to information. The only entities (to date) capable of sense making are sentient beings and not machines.

This being said, ICT has the ability to process data and present subsets of that data that are useful in contexts and are able to aid in decision-making. Machines are also able to act autonomously where they have sufficient means of sensing their environment and have a set of rules or models available to them that enable them to operate (in other words control systems). However those machines when taken out of their operating context would not have the ability to adapt as humans can.

This lack of discriminating between selected sets of data that have relevance and value in a given context and wider pools of unrelated facts suggests there may be a useful intermediate step between data and information that help us understand the

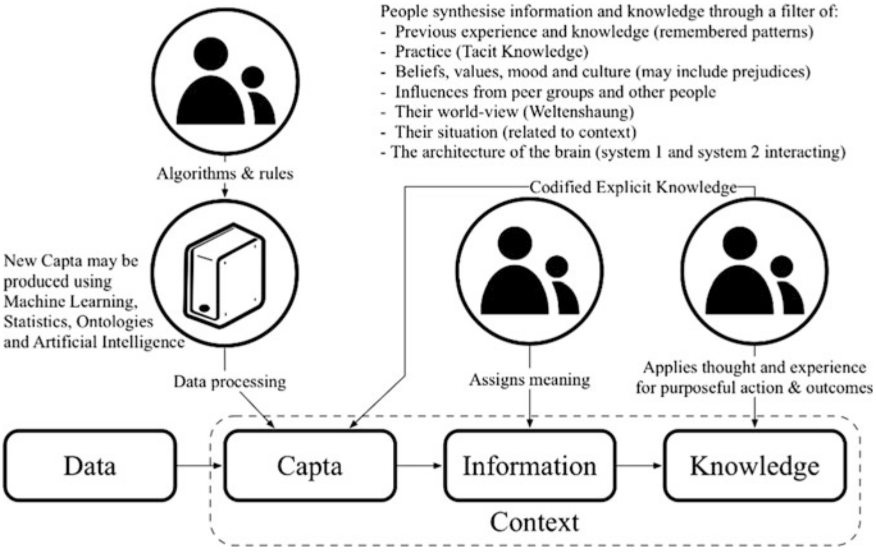


Fig. 6.1 The relationships between data, information, knowledge and the role of ICT and people

true role of ICT in an information or knowledge system. This idea was first suggested by Checkland and Howell [3] where they named these subsets of relevant data, calling them Capta.

The following model is derived from the DIKW model Ackoff [1] first presented. The model is offered as a basic set of defining principles that is useful in helping to architect a useful knowledge facilitation system. The fundamental principle here is that information and knowledge only exists in the grey matter between people's ears. Once this is appreciated it has fundamental ramifications on designing knowledge facilitation systems (Fig. 6.1).

6.4 The Workings of the Human Mind, and Implication on Knowledge Facilitation

Our minds also work in a fundamental ways that are described in Kahneman's [8] book *Thinking Fast and Slow*. Kahanen won a Nobel Prize in behavioural economics for the work he describes in the book. He uses a systems approach that artificially divides the brain into two parts he calls system 1 and system 2. System 1 is the intuitive part of our brain, hardcoded to react very quickly to stimuli (sensing data) without the need to consciously think. This is an evolutionary necessity for survival, initiating rapid reactions to danger to keep us alive. System 1 has much synergy with tacit knowledge in that practice and experience results in actions where we no longer need to use conscious effort to execute them. An example of this is learning to drive a car, where effective use of the car controls is learned through repeated practice, before becoming subconscious (i.e. tacit knowledge). We have all experienced driving home, without conscious memory of the journey when we arrive at our destination. System 2 is rational and lazy, it involves conscious rationalisation of data that costs a lot of personal effort and energy to sustain (to remain focused and concentrated) and compared to system 1, is slow acting. The lazy aspect of system 2 is based on our brain using associative memory that summarises patterns of expected behaviour. An example of this is where we see situations, and our brain simulates and anticipates what will happen next. Optical illusions use these phenomena. There seems to be synergy with system 2 and the concepts of explicit knowledge, where thinking can be discussed and rationalised. Kahneman discusses the interaction of system 1 and 2 in a number of 'heuristics', that describe how decision-making is made and how our judgement may be influenced.

Some of the heuristics discussed include:

- The anchoring effect: This is a subconscious association of being influenced by previously observed quantities. If a person was asked to estimate the price of a bottle of wine, they would give a lower estimate if they had previously been shown the number 5, compared with a higher estimate after being shown the number 30.

- Loss aversion: People are prone to put more effort into avoid losses, than to achieve gains, if we ask a group of people to first choose an option where there is 90% chance of success, and then later choose an option of 10% chance of loss, it is likely the number first choices will be more than the second choices.
- Trusting expert intuition: Intuition is immediate and often subconscious pattern recognition (of past experiences). Experts often have beliefs that are supported by thoughts that easily come to mind when a decision must be made, with no contradictions or competing options. The chance of error is increased when the environment in which the current decision is being made differs from the environments the expert gained their experience and that the expert has had limited practice.

These biases and influences in the way our minds work should be considered in the way Capta is presented, as part of User Experience (UX) design discussed below. UX may be inadvertently misapplied in delivering emotional impact that might prejudice effective decision-making.

6.5 Problems in the Exchange of Codified Explicit Knowledge

During the First World War this communication ([HSTC] [7]) was received by a headquarters from a front line unit in the trenches. The message was relayed by a series of runners

> Send three and four pence, we are going to a dance.

Three and four pence is an example of the British monetary system before decimalisation. It means three shillings and four (old) pence. The original message sent from the trench was:

> Send reinforcements we are going to advance.

This incident shows how a simple and clear message, that has obvious battlefield context that one may have assumed would have increased the chances of an intelligible message being received. This is an example of how codified knowledge may be misinterpreted through a series of senders and receivers to end up as garbage at the final receiver. What possible relevance had petty change for a dance have in the trenches and killing fields of Flanders? Admittedly, the military discipline and attitudes of the time would have suppressed ordinary soldiers questioning the message contents, but they would have also realised and been taught the importance of accuracy to avoid the dire consequences of mistakes in a battlefield context. This is an example that shows us that codified knowledge may not be interpreted as an author intends.

In the corpus of literature on knowledge management, there is surprisingly little written about the effectiveness and efficiency of exchanging and sharing codified explicit knowledge Gourlay [5]. It is generally assumed that once knowledge is expressed and codified that a person receiving the codified knowledge will make sense of it and interpret it in the same way as the author intended? Common sense and our real world experiences are contrary. The vendors of knowledge management systems, that are essentially data storage and retrieval systems provide simplistic systems that do not address the issue of meaning and interpretation of the message, these are left to the user community to determine.

There are new disciplines developing as a result of the continuing digital and information revolution, examples of these include data science and the development of UX or user experience based design. In the UX domain, software applications (UX is not limited to software) are developed to have superior user friendliness and effectiveness, the older ideas about usability centred on being intuitive and simple to navigate. These concepts are extended in UX—the goal is to achieve a positive emotional impact on the user, the use of the software needs to be pleasurable and a joy. This change in thinking also has great implications on how we should be sharing knowledge, the emotional impact of a receiver of the data has influence on the way they interpret it.

In the design of a knowledge facilitation system, one must consider that codified Capta may not be interpreted as intended. This may be alleviated by UX, and how the richness of data. Image, video and audio messages may convey intended meaning more effectively. A powerful human tacit skill is interpreting body language as an integral part of communication. People often misinterpret the emotional intent in e-mail and can easily take offence where none was intended.

6.6 Cultural Aspects that Support Knowledge Facilitation

The cultural and human aspects of knowledge facilitation are vitally important to address, the following questions are pertinent and must be addressed

1. Why would an expert share their knowledge? An expert may well regard that their political and positional power inside an organisation depends on their retaining their expert knowledge. Knowledge is power. Why would they willingly share?
2. Why should a worker use other's knowledge or insights derived from data? Why would someone trust another's knowledge and if they acted on it and things did not work out would blame still be attached to them?

This indicates that for KM to be successful an organisational culture along with its politics and structure needs to be designed to facilitate the use and sharing of knowledge. Knowledge sharing by experts needs to be recognised and rewarded. The reward structure may also extend beyond money, peer group recognition and respect is an important motivator to experts.

Blame culture also suppresses knowledge re-use; mistakes are only made if lessons are not learned from errors, and that errors can be reflected on openly and constructively free of blame apportionment. The culture should enable open admission of mistakes in a spirit of learning from them. In some cultures the loss-of-face in admitting mistakes or isolating an individual is extremely humiliating and threatening, and so local cultures or international teams should take these factors into account.

Leadership is vital to ensure the culture of an organisation is appropriate for knowledge facilitation, leadership must be seem to lead by example and be open to exploring their own mistakes. There is no such thing as an infallible person, no matter how senior they are.

6.7 Technical Aspects that Support Knowledge Facilitation

The advances in ICT with the emergence of 'big data' and the continuing development of the Internet of Things (IoT) are transformative in terms of exploiting data.

A breakthrough in limitations in the variety of data able to be processed has been achieved, allowing unstructured data to be used. The richness of data has also improved with image, audio and video being able to be exploited.

In memory big data technology also enables data to be processed in near real time, where there may be benefits accrued in agile responses to events in a timely manner.

IoT promises the proliferation of cheap ultra-low-power miniaturised wireless enabled smart sensors that reduce the cost of instrumenting ever smaller assets, and connecting these to the internet. This means a transformative step change in the volume of data we may gather and have access to that will provide unprecedented opportunities to understand product lifecycles in minute detail. A useful roadmap toward exploiting IoT is using big data technology, so that the step increase in data can be analysed.

6.8 So What—How Do We Go About Specifying a KF System?

Knowledge facilitation can only work optimally if cultural, socio, political and technical aspects are considered. The technical aspects of knowledge-facilitation are likely to be the easiest to deal with of the four. An organisation that wholly delegates the implementation or running of a knowledge facilitation system to the IT department is making a mistake. A knowledge facilitation system serves people, the right organisational culture, attitudes and behaviours are far more important than technology.

Any developed system must have a purpose and deliver benefits. In the context of a full product lifecycle approach, the agreement and incentives for all organisations to participate and reach a consensus needs to be reached. This may best be achieved within whole industries leveraging standards groups, government or internationally backed research groups.

New standards are emerging that require greater exchange of information, with appropriate commercial safeguards for IP, between manufacturers (OEMs) and operators. An example is The American petroleum Institute's 691 standard (API 691), based on a full product lifecycle perspective on Risk Based (rotating) machinery management. This stipulates Failure Modes and effects (FMEA) type data be provided from manufacturers (OEM) to operator customers for machinery that will be integrated into oil and gas plants, in exchange for the operators allowing OEM access in receiving in-service data from these machines, to enable them to improve their product design.

Another example of standards work in industry wide collaboration in sharing data is in Aerospace, with the SAE HM1 committee and their Integrated Vehicle Health Management (IVHM) initiatives, where airline operators, OEMs, regulators and maintenance providers are working out how IVHM predictive maintenance may be safety justified, in gaining 'maintenance credits' so that predictive maintenance can replace traditional planned maintenance.

The effort must be business lead, with a value proposition that clearly indicates the value in facilitating knowledge throughout the product lifecycle.

Historically, and in the author's personal experience of over 40 years working experience in engineering, there is a schism between those who design, build, operate and maintain complex assets. Often machinery is not used as the designer assumed or intended, and the designer does not fully know how machinery behaves when it is operated. Feeding appropriate Capta and allowing people to work in many of the different phases of the product lifecycle will break down barriers.

Cultural aspects: some cultures are more deferential to elders or to seniors within a hierarchical structure. Other cultures may depend on a whole group conforming and agreeing (embodied in Kaizen) whilst other cultures are more individualistic. This paper makes no assertions about the best type of culture to have, as this may be a value judgement. The design of a knowledge facilitation system must take account of and fit the dominant cultures it serves.

Political aspects. Knowledge facilitation implies sharing relatively sensitive data between departments and separate entities that may be threatening to many people. This threatens the success of a knowledge facilitation system as its implementation or running may be resisted. Lifecycle management requires knowledge to be shared over the whole lifecycle of a product. This will involve commercial and intellectual property concerns to ensure individual organisations are not disadvantaged by

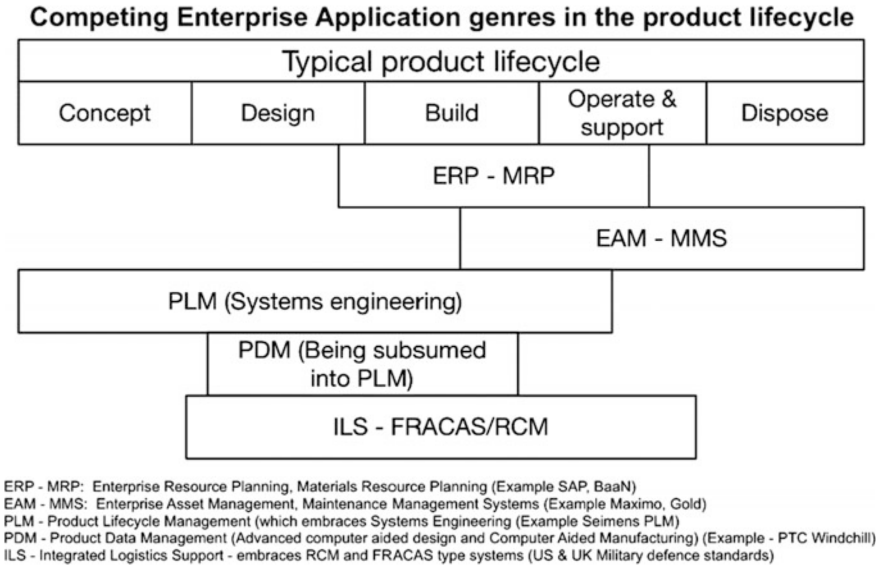


Fig. 6.2 Competing enterprise application genres in the product lifecycle

knowledge sharing. Commercial organisations will need to develop ways of sharing valuable data for the overall benefit of their industries.

Socio aspects: The scenario of incentivising hoarding expert knowledge for power needs to be relieved by making it advantageous to both share and use knowledge. Mentoring, coaching and leadership play vital roles in fostering knowledge sharing. The recognition of Tacit knowledge emphasizes preferment of on the job training and experience, where knowledge transfer will also be more effective if work is openly reflected on. Mistakes should be tolerated to enable learning (Fig. 6.2).

ERP and EAM systems do not embrace knowledge management except that many systems are rolling out big data technology that enables data mining. Examples include SAP HANA that provide in memory data warehousing capability to derive business intelligence.

A good knowledge facilitation system could be built around organisational processes, bringing processes alive, and attracting people into using tools and assistance geared to traversing processes in the most effective and efficient way. Links to data, past examples, both good and not so good, experts (who should be willing and rewarded for helping) may all be integrated into rich media intranet that actively attracts people to use it. The latest thinking and practice from UX (user experience design) should go beyond the usual ‘it should be intuitive to use’ to it will have a positive emotional impact and be a joy to use.

References

1. Ackoff R (1989) From data to wisdom. J Appl Syst Anal 16:3–9
2. API 691 (2016) Risk based machinery management. http://ballots.api.org/cre/some/ballots/docs/B3600_API_691_1st_Edition_Draft.pdf. Accessed 10 May 2016
3. Checkland P, Howell S (1997) Information, systems and information systems: making sense of the field. Wiley
4. Gourley S (2006) Conceptualizing knowledge creation: a critique of Nonaka's theory. J Manag Stud 43(7):1415–1436
5. Gourlay S (2007) An activity centered framework for knowledge management. In: McInerney CR, Day RE (eds) Rethinking knowledge management: from knowledge management to knowledge processes. Information science and knowledge management. Springer, Berlin, Germany, pp. 21–63. ISSN (print): 1568-1300 ISBN: 9783540710103
6. Grant K (2011) Knowledge management, an enduring but confusing fashion. Electron J Knowl Manag 9(2):117–131. http://www.ejkm.com/issue/download.html?idArticle=285
7. [HSTC] 1914 October 15, The temperance caterer, altered in transit, Page 78, Column 2, NA Page 16, London, Middlesex. (NewspaperArchive)
8. Kahnanen D (2011) Thinking fast and slow. Farrar, Straus and Giroux
9. Liebowitz J (2016) Some misconceptions about knowledge management. https://www.k4health.org/blog/post/some-misconceptions-about-knowledge-management Accessed 20 Feb 2016
10. Oxford English Dictionary. http://www.oed.com/viewdictionaryentry/Entry/95568 Accessed 19 Feb 2016
11. Polanyi M (1958, corrected 1962) Personal knowledge, towards a post-critical philosophy. Routledge, London
12. Tsoukas H (2003) Do we really understand tacit knowledge? In: The Blackwell handbook of organizational knowledge management, pp 410–27. doi:10.1.1.18.8864

Part III
The Role of Data, Diagnostics and Prognostics in Through-Life Engineering Services

Chapter 7
Predictive Big Data Analytics and Cyber Physical Systems for TES Systems

Jay Lee, Chao Jin and Zongchang Liu

Abstract In today's competitive business environment, companies are facing challenges in dealing with big data issues for rapid decision making for improved productivity. Many manufacturing systems are not ready to manage big data due to the lack of smart analytics tools. U.S. has been driving the Cyber Physical Systems (CPS), Industrial Internet to advance future manufacturing. Germany is leading a transformation toward 4th Generation Industrial Revolution (Industry 4.0) based on Cyber-Physical Production System (CPPS). It is clear that as more predictive analytics software and embedded IoT are integrated in industrial products and systems, predictive technologies can further intertwine intelligent algorithms with electronics and tether-free intelligence to predict product performance degradation and autonomously manage and optimize product service needs. The book chapter will address the trends of predictive big data analytics and CPS for future industrial TES systems. First, industrial big data issues in TES will be addressed. Second, predictive analytics and Cyber-Physical System (CPS) enabled product manufacturing and services will be introduced. Third, advanced predictive analytics technologies for smart maintenance and TES with case studies will be presented. Finally, future trends of digital twin industrial systems will be presented.

7.1 Introduction

In today's competitive business environment, companies are facing challenges in dealing with big data issues for rapid decision making for improved productivity. Many manufacturing systems are not ready to manage big data due to the lack of smart analytics tools. U.S. has been driving the development Cyber-Physical Systems (CPS) and Industrial Internet to advance future manufacturing. For instance, GE has announced Predix™ as a cloud-based service platform to enable

---

J. Lee (✉) · C. Jin · Z. Liu
NSF Industry/University Cooperative Research Center on Intelligent Maintenance Systems,
University of Cincinnati, 2600 Clifton Ave, Cincinnati, OH 45221, USA
e-mail: jay.lee@uc.edu

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_7

industrial-scale analytics for management of asset performance and optimization of operations [1]. Also, National Instruments introduced Big Analog Data™ three-tier architecture solution [2], as well as LabVIEW Watchdog Agent™ Toolkit to support smart analytics solutions throughout different big data applications [3, 4]. At the same time, Germany is leading a transformation toward 4th Generation Industrial Revolution (Industry 4.0) based on Cyber-Physical Production System (CPPS). It is clear that as more predictive analytics software and embedded Internet of Things (IoT) are integrated in industrial products and systems, predictive technologies can further intertwine intelligent algorithms with electronics and tether-free intelligence to predict product performance degradation and autonomously manage and optimize product service needs,

This book chapter will address the trends of predictive big data analytics and CPS for future industrial Through-life Engineering Services (TES) systems. First, industrial big data issues in TES will be addressed. Second, predictive analytics and CPS enabled product manufacturing and services will be introduced. Third, advanced predictive analytics technologies for smart maintenance and TES with case studies will be presented. Finally, future trends of digital twin industrial systems will be presented.

7.2 Industrial Big Data Issues in TES Systems

Through-life Engineering Services (TES) systems aim at addressing the support requirements for performance-based contracts, of which maintenance is the major engineering service [5]. This has been motivating the users to become increasingly interested in minimizing the whole lifecycle ownership of assets [6]. With the prevalence of smart sensors and IoT technologies such as RFID and MTConnect, data acquisition becomes more and more cost-effective and pervasive, but the question remains if these data will provide us the right information for the right purpose at the right time. Merely connecting sensors to machines or connecting machines to machines will not facilitate rapid decision making. Current manufacturing systems will necessitate a deeper analysis of various data from machines and processes.

The aforementioned issues in TES systems can be addressed by industrial big data. Industrial big data is a systematic methodology to convert different sources of data (sensors, controllers, history, human, fleet peer-to-peer system) into smart actionable information in order to reduce costs and generate business revenues. Industrial big data analytics draws actionable information from raw data collected from various sources to support rapid decision making, so that businesses will be able to increase operation efficiency, improve services, create novel business models, and ultimately, generate more revenues [7]. A research conducted by Accenture and General Electric forecasted that the values created by Industrial Internet of Things and industrial big data could be worth $500 billion by 2020 [7].

The concept of industrial big data in industry is related to big data in information technology, but there are certainly distinctive characteristics between them. Both industrial big data and big data refer to data generated in high volume, high variety, and high velocity ("3 V" problems) that require new technologies of processing to enable better decision making, knowledge discovery and process optimization [8]. Sometimes, the feature of veracity is also added to emphasize the quality and integrity of the data [9]. However, for industrial big data, there should be two additional "V's". One is "Visibility", which refers to the discovery of unexpected insights of the existing assets and/or processes and in this way transferring invisible knowledge to visible values. The other "V" is "Value", which put an emphasis on the objective of industrial big data analytics—creating values. This characteristic also implies that, due to the risks and impacts industry might face, the requirements for analytical accuracy in industrial big data is much higher than big data analytics in general, such as social media and customer behavior [10–13].

Compared to big data in general, industrial big data is usually more structured, more correlated, and more orderly in time and more ready for analytics [10]. This is because industrial big data is generated by automated equipment and processes, where the environment and operations are more controlled and human involvement is reduced to minimum. Nevertheless, the values in industrial big data will not reveal themselves after connectivity is realized by IoT. Even though machines are more networked, industrial big data usually possess the characteristics of "3B" [10], namely:

- Below-Surface

General big data analytics often focuses on the mining of relationships and capturing the phenomena. Yet industrial big data analytics is more interested in finding the physical root cause behind features extracted from the phenomena. This means effective industrial big data analytics will require more domain know-how than general big data analytics.

- Broken

Compared to big data analytics, industrial big data analytics favors the "completeness" of data over the "volume" of the data, which means that in order to construct an accurate data-driven analytical system, it is necessary to prepare data from different working conditions. Due to communication issues and multiple sources, data from the system might be discrete and un-synchronized. That is why pre-processing is an important procedure before actually analyzing the data to make sure that the data are complete, continuous and synchronized.

- Bad-Quality

The focus of big data analytics is mining and discovering, which means that the volume of the data might compensate the low-quality of the data. However, for industrial big data, since variables usually possess clear physical meanings, data integrity is of vital importance to the development of the analytical system.

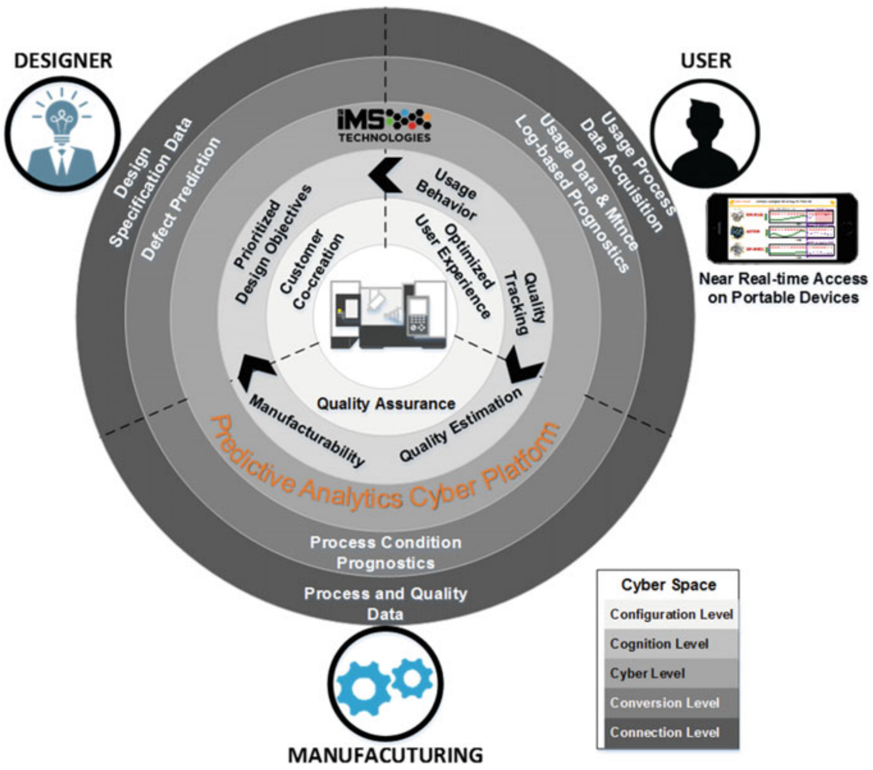


Fig. 7.1 Vision for predictive analytics and Cyber-Physical Systems-Enabled TES system

Low-quality data or incorrect recordings will alter the relationship between different variables and will have a catastrophic impact on the estimation accuracy.

Therefore, simply transferring the techniques developed for general-purpose big data analytics might not work well for industrial big data analytics. Industrial big data requires deeper domain knowledge, clear definitions of analytical system functions, and the right timing of delivering extracted insights to the right personnel to support wiser decision making [10, 14]. Predictive analytics and Cyber-Physical Systems are two core technologies that will help generate the most values from industrial big data in TES systems, which will be introduced in later sections. As Fig. 7.1 shows, predictive big data analytics and Cyber-Physical Systems will not only benefit users from predictive maintenance, but will also nurture customer co-creation through feedbacks to design and proactive quality assurance during manufacturing, which will lead to a more comprehensive TES system.

7.3 Predictive Analytics and Cyber-Physical System-Enabled Manufacturing and Services

Predictive analytics and Cyber-Physical Systems (CPS) are the core technologies of industrial big data [10, 11]. CPS systems require seamless integration between computational models and physical components [15]. Each physical component and machine will have a Digital Twin model in the cyber space composed of data generated from sensor networks and manual inputs. As shown in Fig. 7.2, a CPS can be constructed by following the “5C” architecture, which serves as a guideline for the development of CPS for industrial applications [16]. Specifically, the “5C” architecture refers to the following levels of work flow:

1. Smart Connection Level: From the machine or component level, the first thing is how to acquire data in a secure, efficient and reliable way. It may include a local agent and a communication protocol for transmitting data from local machine systems to a remote central server. Previous research has investigated robust factory network schemes based on well-known tether-free communication methods, including ZigBee, Bluetooth, Wi-Fi, UWB, etc. [17–19].
2. Data-to-Information Conversion Level: In an industrial environment, data may come from different resources, including controllers, sensors, manufacturing systems (ERP, MES, SCM and CRM system), maintenance records. These data or signals represent the condition of the monitored machine systems. However, they must be converted into meaningful information for a real-world application, including health assessment and fault diagnostics.
3. Cyber Level: Once we can harvest information from machine systems, how to utilize it is the next challenge. The information extracted from the monitored system may represent system conditions at that time point. If it can be compared

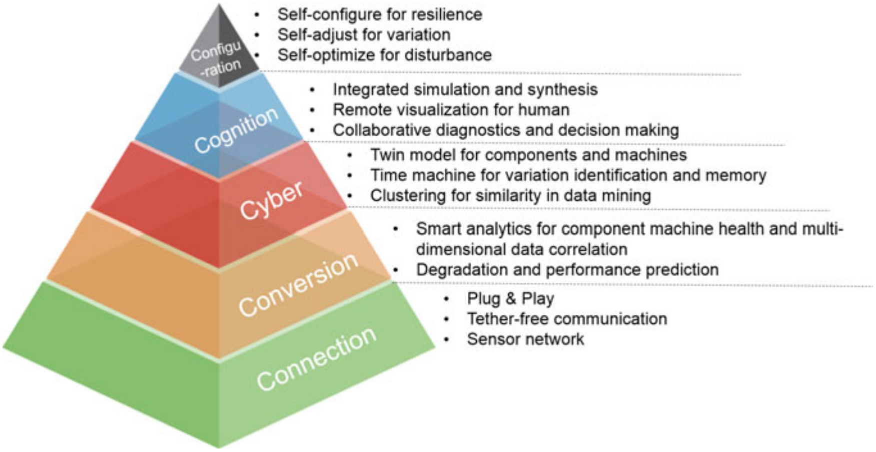


Fig. 7.2 “5C” architecture for Cyber-Physical Systems [16]

with other similar machines or with machines in different time histories, users can gain more insights on the system variation and life prediction. It is called cyber level because the information is utilized in creating cyber avatars for physical machines and building a great knowledge base for each machine system.

4. Cognition Level: By implementing previous levels of CPS, it can provide the solutions to convert the machine signals to health information and also compare with other instances. In cognition level, the machine itself should take advantage of this online monitoring system to diagnose its potential failure and alert its potential degradation in advance. Based on the adaptive learning from the historical health evaluation, the system then can utilize specific prediction algorithms to predict the potential failure and estimate the time to reach certain level of failures.
5. Configuration Level: Since the machine can online track its health condition, the CPS can provide early failure detection and send health monitoring information to operation level. This maintenance information can be fed back to factory systems so that the operators and factory managers can make the right decision based on the maintenance information. At the same time, the machine itself can adjust its working load or manufacturing schedule in order to reduce the loss caused by the machine malfunction and eventually achieve a resilient system.

For the first level, communication protocols play a significant role to enable tether-free communication between machines and data acquisition systems. In this way, recently developed communication protocols such as MTConnect [17] can help users acquire controller signals. Authors of the research in [18] demonstrated that using new communication protocols, in this case MTConnect, paves the way of acquiring data from band saw machines and other manufacturing equipment. Although these methods are helping to make data acquisition more efficient, the challenge of dealing with different sources of data is still in place [19].

The second level of the architecture, data to information conversion, has also received considerable attention specifically for prognostics and health management (PHM). Lee et al. [20] provides a relatively comprehensive review on current PHM approaches for rotary machinery. Trendafilova et al. [21] presented a non-linear data analysis method using accelerometer measurements to identify the backlash severity for industrial robot joints. Liao et al. [22] developed a method to use multiple baselines for identifying faults in band saw machines axis. They used vibration, temperature and torque measurements to train various baselines in self-organizing maps models. As it is obvious with these cases, using induced faults and laboratory situation can cover some key failure signatures of the target system but it is not possible to identify all the possible failure modes which happen in real-life situation.

The third level of the proposed "5C" architecture, the cyber level, intends to provide more intelligent and time-based methods. Using equipment history and algorithms that improve and adapt themselves provides more reliable and robust methods for equipment health monitoring and life estimation. Such algorithms are able to learn the equipment behavior over time and improve initial failure signatures. Even if the original target equipment is unavailable or failed, these models will still be available and can be applied to any other similar equipment. These cyber-models use historical data to improve themselves and data from similar machines to gain more knowledge. Wang et al. [23], developed a trajectory similarity based prognostics method which uses historical data to identify the remaining useful life of assets. Yang et al. [24] developed an adaptive prognostics and health management method for adaptively identifying new working regimes in the data and build new models base on them. Lapira [25] developed a method to use different clustering algorithms to perform machine to machine comparison and generate the health status of wind turbines and industrial robots. This fleet based similarity approach provided more accurate estimation of machines' status by peer comparison and identifying a more reliable baseline.

The cognition level intends to apply decision making and reasoning methods to recommend actionable operations to maintain optimal production while extend the lifetime of assets. There are few researches that focus on using real machine status for decision making such as the work by Haddad et al. [26] where authors used the asset remaining useful life (RUL) as input to option theory to identify the appropriate time for maintenance actions. This study only focused on deciding the maintenance time and did not consider changing working regime and reducing load of the machine as possible options.

The configuration level provides machine with self-adjusting and self-configuration capability. Most of the current research is focused on keeping humans in the decision making loop. Therefore, there are significant research opportunities on developing the concept of self-configuration and self-adjustment concept.

The "5C" architecture has indicated that Cyber-Physical Systems is focused on transferring raw data to actionable information, understanding process insights, and eventually improve the process by evidence-based decision making. Improved processes will further increase productivity and reduce costs. This aligns with the mission of TES systems, which supports usage performance requirements throughout product lifecycle and create values for customers.

7.4 Advanced Analytics for Smart Maintenance and TES with Case Studies

In industrial applications, the "5C" architecture can be applied hierarchically to different levels including components, machines, and fleets. At each level, specific analytics are required to generate useful information from raw data and consequently discover useful knowledge about the system.

1. Component level: At this level, digital twins from critical components of each machine are modeled in the cyber space. They work in parallel with the physical component while possessing huge differences: they are not bounded by time or location. These digital twins capture significant changes in the health status of each component and once the physical component is degraded, they will give prediction of the remaining useful life. Additionally, as these digital twins are on the cloud, they can interact with other components even though they are geographically distributed. Such models are very inclusive as they log lifespan of components undergoing various stress level and working regimes. Consequently, the system will gain self-awareness.
2. Machine Level: This level incorporates knowledge generated in the component level in addition to machine operation history, system settings and other attributes to create a digital twin for each machine. Adequate analytical methods have to be applied at data-to-information Conversion Level and Cyber Level to generate machine level performance and health metrics. At this stage, digital twins of similar machines are comparable to each other to identify low performance machines regardless of working regime.
3. Fleet Level: As mentioned before, cyber models are not bounded by time or location. This advantage provides opportunity to design and incorporate methods for reactively modifying the production flow. For example, leveraging the historical machine performance data and component status (from component and machine levels), it is possible to optimize the working regime among the fleet in order to maximize the life span of all components and at the same time maintain optimal productivity. This level brings self-maintainability and self-configurability to the system.

7.4.1 Advanced Predictive Analytics and Algorithms

The effectiveness of the proposed CPS architecture relies on the performance of the data analysis and management functions deployed in the cyber level. Served as a bridge connecting the lower level data acquisition and upper level cognition functions, the cyber level is required to autonomously summarize, learn and accumulate system knowledge based on data collected from a group of machines. The system knowledge includes possible working regimes, machine conditions, failure modes and degradation patterns, which is further used by cognition and reconfiguration functions for optimization and failure avoidance. On the other hand, because of complications in machine configuration and usage patterns, autonomous data processing and machine learning is of high priority since the traditional ad hoc algorithm model can hardly be applied to complex or even unexpected situations. New methods have to be developed to perform these tasks and generate appropriate results. In this section, we introduce the “Time Machine” as the framework for performing analytics on the cyber level. This framework consists of three major parts [27]:

1. Snapshot Collection: Information is continuously being pushed to the cyber space from machines. The role of snapshot collection is to manage the incoming data and store the information in an efficient fashion. Basically, to reduce required disk space and process power, snapshots of machine performance, utilization history and maintenance have to be recorded instead of the whole time-series. These snapshots are only taken once a significant change has been made to the status of the monitored machine. The change can be defined as dramatic variation in machine health value, a maintenance action or a change in the working regime. During the life cycle of a machine, these snapshots will be accumulated and used to construct the time-machine history of the particular asset. This active time-machine record will be used for peer-to-peer comparison between assets. Once the asset is failed or replaced, its relative time-machine record will change status from active to historical and will be used as similarity identification and synthesis reference.
2. Similarity Identification: In cyber level, due to availability of information from several machines, the likelihood of capturing certain failure modes in a shorter time frame is higher. Therefore, the similarity identification section has to look back in historical time machine records to calculate the similarity of current machine behavior with former assets utilization and health. At this stage, different algorithms can be utilized to perform pattern matching such as match matrix [28], fleet-based fault detection [25], and trajectory similarity-based methods [23]. Once the patterns are matched, future behavior of the monitored system can be predicted more accurately.
3. Synthesis Optimized Future Steps: Predicting remaining useful life of assets helps to maintain just-in-time maintenance strategy in manufacturing plants. In addition, life prediction along with historical time machine records can be used to improve the asset utilization efficiency based on its current health status. Historical utilization patterns of similar asset at various health stages provide required information to simulate possible future utilization scenarios and their outcome for the target asset. Among those scenarios, the most efficient and yet productive utilization pattern can be implemented for the target asset.

New algorithms have to be designed to comply with the proposed Time Machine framework. In this section, two representative machine learning and knowledge extraction methodologies are introduced for performing health assessment and prognostics within the CPS structure.

7.4.1.1 Similarity-Based Fleet-Sourced Health Monitoring

Considering a machine fleet, similarity always exists among machines—machines that are performing similar tasks or at similar service time may have similar performance and health condition. Based on such similarity, machine clusters can be built, as a knowledge base representing different machine performance and working conditions.

As for algorithms, unsupervised learning algorithms such as Self-Organizing Map (SOM) and Gaussian Mixture Model (GMM) can be used for autonomously creating clusters for different working regime and machine conditions. The adaptive clustering methodology [24] utilizes an on-line update mechanism: the algorithm compares the latest input (Time Machine) to the existing cluster and tries to identify one cluster that is most similar to the input sample using multidimensional distance measurement. Search of similar cluster can end with two results: (1) Similar cluster found. If it is this case, then the machine from which the sample has been collected will be labeled as having the health condition defined by the identified cluster. Meanwhile, depending on deviation between existing cluster and the latest sample, the algorithm will update the existing cluster using new information from the latest sample. (2) No similar cluster found. In this case, the algorithm will hold its operation with the current sample until it sees enough count of out-of-cluster samples. When number of out-of-cluster samples exceeds a certain amount, it means that there exists a new behavior of the machine that has not been modeled so that the algorithm will automatically create a new cluster to represent such new behavior. In such case the clustering algorithm can be very adaptive to new conditions. Moreover, the self-growing cluster will be used as the knowledge base for health assessment in the proposed cyber space. With such mechanism, different machine performance behaviors can be accumulated in the knowledge base and utilized for future health assessment.

7.4.1.2 Prognostics of Machine Health Under Complex Working Conditions

After the health condition and the working regimes are identified for each machine, the next step is to predict the remaining useful life (RUL). First, using utilization history and measurement data, the relationship between machine degradation and the utilization (stress) history is built. Many existing prediction algorithms fail to perform well for in-field machines because they cannot handle dynamic or complex-working regimes that may alter the actual degradation path from previously learned ones. The proposed utilization matrix based prediction is grounded on the understanding that the fundamental reason for machine degradation is not only time, but also other stress factors. As a consequence, a general-purpose prediction algorithm has to be based on the stress versus life relationship.

For systems such as CNC machines, more dimensions (e.g. material hardness, machining parameters, volume of removed material, etc.) need to be added to the stress matrix to cover all major factors that cause degradation. After the definition of stress matrix, machine learning algorithms such as Bayesian Belief Network (BBN) and Hidden Markov Model (HMM) can be used to relate the different degradation rates observed in machine fleet to corresponding stress history.

Eventually degradation rate can be generated for prediction under different usage patterns that may occur in real world applications.

7.4.2 Case Studies

7.4.2.1 Self-aware Band Saw Machine

The core components of band saw machines are band saws used for cutting. As cutting volume grows, the band saws will gradually wear down, which results in a decline in processing efficiency and quality. For this reason, the plant must arrange a large number of workers to keep an eye on the machine operations and the wear of band saws, and determine the replacement timing based on experience. As quality requirements vary with different cutting tasks, and factors influencing quality cannot be root caused easily, the healthy band saws would be replaced well before they break. Thus, it is necessary to gather processing data from band saw machine controllers and add-on sensors, and develop a predictive CPS platform for band saw degradation analysis and prognostics, thereby making the band saw machine more intelligent by providing customers with visualized productivity management services.

In the course of processing, an intelligent band saw machine can analyze data in near real time: It first identifies condition parameters of the current work piece, and then extracts diagnostic characteristics from vibration signals and other critical parameters. After normalizing diagnostic characteristics in light of working conditions, it maps the current diagnostic characteristics to areas on the health map representing the current health stage. Such information is divided into three categories: working condition information, diagnostic information, and health status information. With voluminous lifecycle information files on band saws, users can conduct data mining through big data analysis.

While making band saw machines self-aware and intelligent, the manufacturer developed an intelligent cloud service platform to provide users with customized band saw machine health and productivity management services. As shown in Fig. 7.3, after status information gathered by band saw machines is transmitted to the cloud for analysis, users can get the health condition of key components, degradation of band saws, operation parameter matching and risk assessment through the user interface on portable devices or web interface. This makes every band saw machine operating condition quantitative and visual. With the platform, users can also manage their production plans, and manage band saw machines and band saws as required by the production tasks. When a band saw is worn to a point that it cannot meet machining quality requirements, the system will automatically remind the user to replace it, and automatically generate an order for the band saw in the material management ERP system. While dramatically boosting the efficiency of human resources, it avoids uncertainties brought about by management based on

Machine	
Health Condition:	0.92
Accu. Power Time:	164 hours
Accu. Main Motor Time:	64 hours
Accu. Pump Motor Time:	26 hours
Accu. Hydraulic Motor Time:	164 hours
Accu. Cutting Time:	6943 hours
Blade	
Current Blade:	ASSAB 83
Health Condition:	0.26
Accu. Tension Time:	3024.5 hours
Accu. Area:	26615 cm^2
Cutting Rate:	6.2094259 cm^2/min

Last cutting record received at 2015-05-04 00:11:25

Fig. 7.3 Intelligent management of band saw machine

experience. Meanwhile, the service life of band saws is prolonged, and quality management is conducted in a quantitative and transparent manner.

After demonstration at the International Manufacturing Technology Show (IMTS) 2014 in Chicago, the manufacturer's intelligent band saw machines and intelligent cloud services drew great attention. Seen as outstanding demonstration of intelligent equipment, such products and services have won great popularity among customers.

7.4.2.2 From Lean to Smart—Production Line Smart Maintenance

Automobile Manufacturer B has introduced a prediction analysis model in the health management of industrial robots. As such industrial robots were widely applied under different working conditions for different manufacturing purposes, installing external sensors for them was not feasible, and their health status should be analyzed based on parameters obtained in the controller. One type of the industrial robots deployed by Automobile Manufacturer B is the six-axis robotic arm, which would completely shut down when a fault occurs on any axis. To address this problem, Automobile Manufacturer B first identified the robotic arm working conditions based on the revolving speed signals of its servo axes and then established a health evaluation model for the status parameters (such as torque and

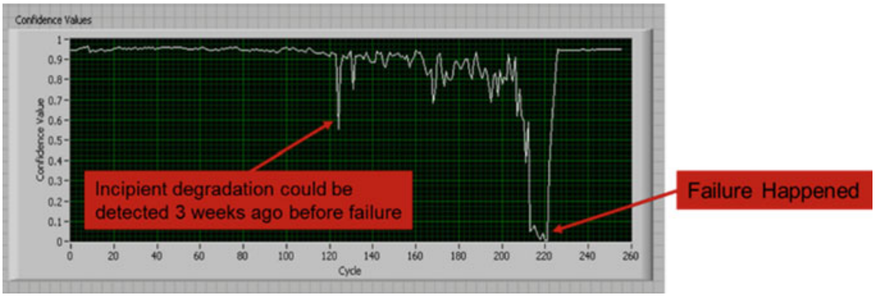


Fig. 7.4 Predictive analysis results for industrial robotic arms

temperature) under each working condition, so as to predict a fault three weeks in advance.

After that, Automobile Manufacturer B started to roll out the prediction analysis model on the six-axis servo robotic arms, conducted cluster analysis based on their types and working conditions, and formed "cyber fleets" for the various types of robotic arms. For each "cyber fleet," Automobile Manufacturer B adopted cluster modeling to analyze data about the covered robotic arms, judged the abnormality of each robotic arm by comparing it with the whole community, and sorted all robotic arms in the community based on their health status (Fig. 7.4).

After conducting quantitative analysis on the health status of the robotic arms, Automobile Manufacturer B adopted an Internet-based model to manage the analysis results, and established an online monitoring system for the "cyber factory". In the "cyber factory", administrators can manage equipment status in a vertical and all-around manner at the levels of production system, production line, work station, single equipment and even key components, and carry out the maintenance and production plans based on the current equipment status. This system can generate a health report every day, which analyzes all equipment in the production line, sorts them based on their health status, and indicates the health risks and problematic parts of all equipment for equipment management personnel. In this way, potential risks will be identified in routine spot inspections and unnecessary inspections and maintenance can be effectively avoided. As a result, Automobile Manufacturer B successfully achieved the transition from preventative to predictive maintenance.

7.5 Future Trends of Digital Twin Industrial Systems

Through the discussion in precious sections, as shown in Fig. 7.5, it is evident that future industrial systems will shift from machine-based to evidence-based decision making, from solving visible problems to avoiding invisible problems [29], and from product-centric quality control to user-centric value creation.

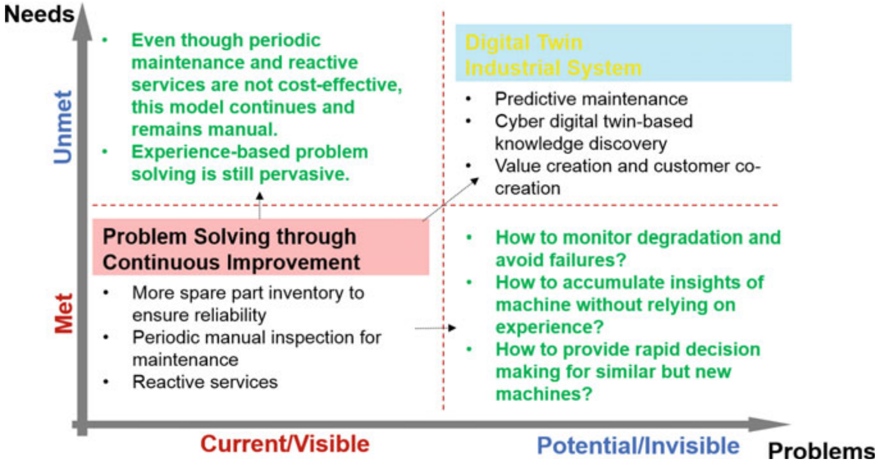


Fig. 7.5 Opportunity space of TES Digital Twin Industrial System

First, transformation from machine-based to evidence-based decision making will rapidly take place. Traditionally, manufacturing system management heavily depends on experienced personnel. In an aging society, knowledge loses with retired workforce. Therefore, a smart analytical system is needed to transform experience-based know-how into evidence-based decision making for sustainable operation.

Then, transformation from solving visible problems to avoiding invisible issues will become a new focus of the industry to change the mindset of smart maintenance. Manufacturing issues can generally be divided into visible and invisible categories [13]. Through smart analytics of interconnected multidimensional systems, correlations and causal functions can be modeled so that meaningful and actionable information can be extracted.

Eventually, transformation from product-centric quality control to user-centric value creation will naturally become the objective of the aforementioned efforts. Product quality is important, but that shall not be the end of TES systems. The final objective of manufacturing products and providing services is to optimize user experience and in return to improve design to further advance product features. Users will drive the needs of both product features and service models, and predictive analytics and CPS technologies will be the foundation of revealing and fulfilling such needs.

In future digital twin industrial systems, data will remain the most important medium to provide customized products and services for users. Through data, customers will be connected with the manufacturing systems closely and be involved in the design, manufacturing, and maintenance phases. Digital Twin Industrial System will not merely become a transformation of manufacturing systems, but a more profound and revolutionary change in business models, service models, supply chains and value chains. Its fundamental motivation comes from innovative technological changes in the business model and intelligent service system.

References

1. General Electric Predix. https://www.ge.com/digital/predix
2. National Instruments Big Analog Data™ Solutions. http://www.ni.com/white-paper/14667/en/
3. Center for Intelligent Maintenance Systems Development of Smart Prognostics Agents (WATCHDOG AGENT®). http://www.imscenter.net/front-page/Resources/WD.pdf
4. National Instruments Watchdog Agent™ Prognostics Toolkit for LabVIEW—IMS Center. http://sine.ni.com/nips/cds/view/p/lang/en/nid/210191
5. Roy R, Shaw A, Erkoyuncu JA, Redding L (2013) Through-life engineering services. Meas Control 46:172–175
6. Roy R, Erkoyuncu JA, Shaw A (2013) Product-service integration for sustainable. Solutions. doi:10.1007/978-3-642-30820-8
7. Millman N (2015) Big data to unlock value from the industrial internet of things. In: ComputerWeekly.com’. http://www.computerweekly.com/opinion/Big-data-to-unlock-value-from-the-Industrial-Internet-of-Things
8. Laney D (2012) The importance of “Big Data”: A definition. Gartner
9. University V What is big data? http://www.villanovau.com/resources/bi/what-is-big-data/#.Vk4vXHarSUk
10. Lee J (2015) Industrial Big Data. Mechanical Industry Press, China
11. Lee J (2014) Keynote presentation: recent advances and transformation direction of PHM. roadmapping work. Meas Sci Progn Heal Manag Smart Manuf Syst Agenda
12. Lee J, Kao H-A, Yang S (2014) Service innovation and smart analytics for industry 4.0 and big data environment. Procedia CIRP 16:3–8
13. Lee J, Lapira E, Bagheri B, Kao H (2013) Recent advances and trends in predictive manufacturing systems in big data environment. Manuf Lett 1:38–41. doi:10.1016/j.mfglet.2013.09.005
14. Courtney B (2014) Industrial big data analytics: the present and future. InTech Mag
15. Foundation NS Program Solicitation: Cyber-Physical Systems (CPS). http://www.nsf.gov/pubs/2015/nsf15541/nsf15541.htm
16. Lee J, Bagheri B, Kao H-A (2015) A Cyber-Physical Systems architecture for industry 4.0-based manufacturing systems. Manuf Lett 3:18–23
17. Vijayaraghavan A, Sobel W, Fox A, et al (2008) Improving machine tool interoperability using standardized interface protocols: MT connect. Lab Manuf Sustain
18. Neelakanta PS, Dighe H (2003) Robust factory wireless communications: a performance appraisal of the Bluetooth and the ZigBee colocated on an industrial floor. In: Industrial electronics society 2003. IECON’03. 29th annual conference IEEE. pp 2381–2386
19. Hung M-H, Cheng F-T, Yeh S-C (2005) Development of a web-services-based e-diagnostics framework for semiconductor manufacturing industry. IEEE Trans Semicond Manuf 18:122–135
20. Lee J, Wu F, Zhao W et al (2014) Prognostics and health management design for rotary machinery systems—reviews, methodology and applications. Mech Syst Signal Process 42:314–334. doi:10.1016/j.ymssp.2013.06.004
21. Trendafilova I, Van Brussel H (2001) Non-linear dynamics tools for the motion analysis and condition monitoring of robot joints. Mech Syst Signal Process 15:1141–1164. doi:10.1006/Mssp.2000.1394

22. Liao L, Lee J (2009) A novel method for machine performance degradation assessment based on fixed cycle features test. J Sound Vib 326:894–908
23. Wang T (2010) Trajectory similarity based prediction for remaining useful life estimation. University of Cincinnati
24. Yang S, Bagheri B, Kao H-A, Lee J (2015) A Unified framework and platform for designing of cloud-based machine health monitoring and manufacturing systems. J Manuf Sci Eng 137:40914
25. Lapira ER (2012) Fault detection in a network of similar machines using clustering approach. Univeristy of Cincinnati
26. Haddad G, Sandborn PA, Pecht MG (2012) An options approach for decision support of systems with prognostic capabilities. IEEE Trans Reliab 61:872–883. doi:10.1109/TR.2012.2220699
27. Lee J, Bagheri B, Kao H-A (2015) A Cyber-Physical Systems architecture for industry 4.0-based manufacturing systems. Manuf Lett 3:18–23
28. Djurdjanovic D, Lee J, Ni J (2003) Watchdog Agent—an infotronics-based prognostics approach for product performance degradation assessment and prediction. Adv Eng Inf 17:109–125. doi:10.1016/j.aei.2004.07.005
29. Lee J, Bagheri B, Jin C (2016) Introduction to cyber manufacturing. Manuf Lett 8:11–15

Chapter 8
Development and Operation of Functional Products: Improving Knowledge on Availability Through Use of Monitoring and Service-Related Data

John Lindström, Elisabeth Källström and Petter Kyösti

Abstract The book chapter addresses which measures five manufacturing companies have taken, or plan to take, regarding use of data originating from monitoring, service, support, maintenance, repairs as well as other sources, in order to improve the knowledge on availability in the context of providing Functional Products. Commonly, the objective of Functional Products is to provide a function to customers with a specified level of availability (or improvement of productivity or efficiency). The results indicate that systematic planning and collection of relevant data, which is either pre-processed on-board (i.e., locally) or sent as is to central or cloud-based storage and processing, in combination with additional necessary data from other sources, is crucial to build knowledge in order to uphold and improve the level of availability agreed upon with customers. As the use of software in Functional Products increases, the knowledge on availability related to software must be augmented—which can be a challenge for many companies whose operations have been rooted in hardware. Further, the results reveal that getting high-quality input is key in order to use the collected data for analytics and to find root causes. The latter may change how the current value-chain operates and secures the quality of necessary data when providing functions to customers if partners are involved in the provider consortium.

J. Lindström (✉)
ProcessIT Innovations R&D Centre, Luleå University of Technology,
971 87 Luleå, Sweden
e-mail: john.lindstrom@ltu.se

E. Källström
Volvo CE, Volvo Construction Equipment, 63185 Eskilstuna, Sweden

P. Kyösti
Computer-Aided Design, Luleå University of Technology, 971 87 Luleå, Sweden

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_8

8.1 Introduction

The book chapter, based on an empirical study involving five companies, addresses improvement of knowledge on availability in the context of Functional Products (FP) by studying which measures the companies have taken, or plan to take, regarding use of data originating from monitoring, service, support, maintenance, repairs as well as other sources. A trend among manufacturing companies is to increasingly incorporate additional soft parts in order to extend their regular product offers. In addition, the providers take on more responsibility and in some cases also the ownership of the product throughout the entire lifecycle of the product. This trend enables the customers to concentrate on their core business, to stay focused on key processes as well as their customers' pain points. Examples of additionally complex and value-adding business models, compared to products and services, are performance- or result-oriented business models which may be used to remain competitive and profitable over time on global markets. FP is an example of such a business model, and the FP concept [1–4] intertwines hardware, software, service-support system and management of operation into a combined effort providing a function to customers with an agreed-upon level of availability or improved productivity or efficiency. Until the very end of the FP lifecycle, operation of the FP must be managed, further improved and optimized, since the intent with FP is to increase the long-term value for both the customer and the provider and thus maintain a sustainable win-win situation [5–7]. The FP concept shares commonalities and similar features with, for instance, Functional Sales (FS) [8], Extended Products [9], Total Care Product (TCP) [1], Product-Service System (PSS) and Industrial Product-Service Systems (IPS^2) [10, 11], Servicizing [12], Service Engineering [13] or Through-life Engineering Services (TES) [14] in terms of increasing the focus on added soft parts such as service components, knowledge, know-how, problem and risk mitigation [11], etc., additionally offered. The FP concept which originates from hardware aspects has most commonalities with PSS/IPS^2, TCP, and TES, however having additional complexity development-wise.

FP, whose customer contracts can range up to 20–30 years, differ from offering the same hardware and software as a product with services. Important differences are that the provider retains the ownership, takes on risks and additional responsibilities from the customer, and co-creates value together with the customer. Further, the provider needs to honour the agreed-upon level of availability (and/or other contract parameters such as improved efficiency or productivity). To be able to honour what has been agreed upon, in an optimal manner, the provider needs to be able to monitor the function. Alonso-Rasgado et al. [1 p 533] formulate this as "*the defined functional outputs may be contractual agreements. Functional product provision must include monitoring of quantitative performance measures that determine whether the level of functional provision reaches, or exceeds, the levels specified in the contractual agreement*". This is necessary in order to be able to predict problems before they occur and act proactively, rather than reactively as a problem or breakdown has happened. FP availability can be seen as a function of reliability and maintainability [15] of the main constituents that are part of delivering the function.

The current research on FP has mostly been directed towards the hardware and service-support system by modelling and simulation of these two constituents (and in particular the reliability and maintainability). For instance, Löfstrand et al. [16, 17] propose an availability simulation framework and Reed et al. [18–20] outlin a language facilitating modelling of the service-support system. Further, Kyösti and Reed [21] present a simulation model whose intended use is as a decision-support tool in FP development situation, including the implications of geographical placement of service-support centres. In addition, a high-level outline of the challenges and need to model and simulate the whole FP has been made by Pavasson et al. [22] and FP availability, legalities, information security, criticality and adequacy, when planning to use cloud services in FP have been assessed by Lindström et al. [23]. Regarding risks, Reim et al. [24] identify ten potential risks that can hinder FP and posit that contractual risks, breakdowns, technical risks and inappropriate organizational structural risk are the most prominent risks for FP. TES concern the technical services that are necessary to guarantee the required and predictable performance of a complex engineering system throughout its expected operational life with the optimum whole life cost [14], and focus on development of technology and engineering solutions to address the new support requirements for these performance-based contracts. Examples of related TES research range from Schwabe et al. [25], who examine how aerospace products may benefit from effective capture, prediction and reduction of risk probability, to Rajagopal et al. [26], who assess the impact (or costs) from software obsolescence in the defense industry. Further, Okoh et al. [27] review recent findings pertaining to the prediction of remaining useful life in order to provide prognostics support. In addition, Datta and Roy [28] note that availability-based contracting has replaced traditional service procurement practices, in particular, within defense procurement, and discuss the enhancement of existing knowledge in cost estimation models with availability-type support service contracts focusing on equipment availability targets and predefined service levels. Such contracts measure the delivery of services and products to be available for use by the customer. Regarding PSS/IPS^2 research, Reim et al. [29] have developed a PSS risk management decision support framework, which can be used to understand the risk and availability implications when offering PSS. Further, in a review of PSS design methodologies, Vasantha et al. [30] distinguished a clear need to define a common ontology for aspects such as characteristics of requirements, product services, stakeholders, the design processes, life-cycle stages, outcomes, business models and support systems. They also note that challenges in the development of sustainable PSS are that sustainability, availability and higher customer satisfaction must be guaranteed over the lifecycle due to intensified service and knowledge content and data sharing throughout the product life-cycles [30]. Lastly, according to Tukker [31], the most important contribution of the post-2006 literature is to the understanding of what PSS development means for a company's structure, culture, capabilities and management. Thus, companies who want to embrace the PSS concept will need to focus more on product availability for clients rather than product production; an emphasis on diversification through services rather than product ranges, and the need for staff to possess both product knowledge and relation management skills [31].

Currently, there is a lack of literature addressing which measures FP providers have taken or plan to take in order to improve knowledge on FP availability. The current literature on FP availability focuses to a large extent on the hardware and service-support system constituents. Thus, the research question addressed in this chapter can be formulated as: which measures to improve knowledge on FP availability are taken or planned by FP providers pertaining to the use of data originating from monitoring, service, support, maintenance, repairs as well as other sources? Thus, the purpose of the chapter is to outline a set of measures improving the knowledge on availability, which can be used by FP providers together with their customers in the manufacturing industry as well as with researchers.

8.2 Methods

The research approach employed in this study has been based on in-depth qualitative studies with six respondents representing five manufacturing companies. The empirical studies were conducted using semi-structured open-ended interviews [32, 33] with respondents working for companies active in the Faste Laboratory at Luleå University of Technology, Sweden, which is a VINNOVA Excellence Centre focusing on FP Innovation. Further, two additional companies, Komatsu Forest (which develops forestry equipment with availability-oriented solutions) and Electrolux (which sells functional or "all-inclusive" offers to customers), were also part of the empirical studies. Thus, the respondents were well aware of and knowledgeable regarding FP. The respondents were professionals responsible for marketing, services, strategy, development and sales at the following five international companies:

- Gestamp Hardtech AB (one respondent—manager tool design and development)
- Volvo CE (one respondent—chief project manager)
- Electrolux (one respondent—regional category manager)
- Komatsu Forest (one respondent—Executive Vice President)
- Bosch Rexroth AB (two respondents—technical product and service managers)

The purpose of having multiple companies with diverse focus was to ensure an advance in the knowledge of improving FP availability through use of monitoring and service-related data in the context of FP, considering the similarities and differences between the companies (cf. [34]). Although the companies have different offerings, they all face the common challenge of how to best develop, market and sell FP and/or similar concepts such as TES or PSS/IPS[2], either as a provider in a partner consortium or as part of their own offerings. The companies are all manufacturing companies with roots in hardware development. However, additional complimentary components have been added to their customer offerings. What the additional components comprise and their weight or importance differs depending on industry and customer segments served. Some of the companies aim to increase their revenue from soft parts; i.e., services, knowledge or know-how, etc., as well as FP sold globally. Thus, the FP planned or currently offered by the companies vary and have different emphasis on the composition of hardware, software, service support system and management of operation.

Initially, semi-structured interviews were used, with open-ended questions [32, 33] allowing the respondents to give detailed answers and the possibility to add extra information where deemed necessary [35]. The duration of the interviews was between one and two hours. In order to reduce response bias, the respondents came from various parts of the organizations as well as different levels i.e., strategic, tactical and operational units. In order to strengthen the validity of the study, data were continuously displayed using a projector during the interviews, allowing the respondents to immediately read and accept the collected data. If immediate reading and acceptance was not possible, the interview transcript was read and accepted afterwards by the respondents. After that, the collected data were displayed and analyzed using matrices (cf. [36]). The analyzed data were finally summarized into a matrix, ordered and sorted according to relation towards hardware, software, service-support system or management of operation, comprising a number of aspects regarding improving knowledge on FP availability through use of monitoring and service-related data. For reasons of confidentiality, only an aggregated view of the analysis is presented.

8.3 Improving Knowledge on Functional Product Availability Through Use of Monitoring and Service-Related Data

In order to structure the data collected, the proposed FP main constituents [2, 4], i.e., hardware, software, service-support system and management of operation, were used to order and sort the data. Below, the main constituents are thus used when outlining the results of the study. The data are further ordered by actual measures taken now, followed by planned measures (actual ones are indicated by a ‘black bullet’, whereas a planned is marked by a ‘hollow bullet’). After the results, additional reflections from the respondents are highlighted and discussed. At the end of this section, an analysis of the results is presented. The analysis mainly considers the results through the lenses of data and analytical methods.

8.3.1 Hardware

The hardware constituent of FP is commonly the oldest constituent and has evolved in many cases from originally having been a product (potentially sold with add-on services). Thus, there is plenty of history, formal knowledge and tacit knowledge (i.e., know-how) on the hardware and its availability—which is a strength when moving towards providing more complex offers such as TES, PSS/IPS[2] or FP. As previously mentioned, the research on FP availability has to a large extent focused on hardware and the service-support system. The following input from the respondents on the hardware is summarized below and further analyzed at the end of this section:

- The hardware is monitored using an aggregated health index and values based on various related parameters measured in order to facilitate predictive/proactive maintenance. The aggregated health index should give an overall view of the system status.
- All critical parts and activities are monitored by sensors. Further, depending on customer requirements, the output from the function is measured in order to meet quality requirements. Possibly, there will be an increase in the number of measurements and parameters measured on the input and output from the function in order to couple the measurement outcomes to input materials, design choices and pre-operations, etc.
- Estimations regarding the expected life-time and reliability, using fault trees, for instance, are made, and simulations are an important part of this.
- Investigations are made to determine other features or aspects of the hardware that can be measured and monitored with sensors and other means to retrieve additional relevant data from the equipment. Today, data concerning running cycles, usage, loads, etc. are stored, as are error codes generated at indications or actual errors/faults. The data are analyzed centrally.
- FTA-analysis and analyses on what breaks down at different intervals, and how this can be improved, are used on a continuous basis. If some machines are run under certain conditions, modifications may be necessary, for example, larger diesel tanks may need to added. Today, we have quite a few sensors within the machines measuring various parameters.
- Design for diagnostics, i.e., to improve possibilities for finding errors and diagnose equipment, is gaining in importance. Further needed is the creation of additional customer value from this.
- There is a need to test additional sub-systems and simulate them in order to save (development/test) time, build knowledge on service need and improve quality.
- In the future, our hardware may be equipped with sensors measuring wear and tear in bearings and other components with consistent wear and tear (e.g., substrate sensors measuring the degree of wear and tear).

The data points towards the necessity to monitor and measure the output from the FP at the same time as monitoring a number of parts and components of the FP in operation. The "sensoring" will in many cases increase as the knowledge on what is possible to do (and willingness to learn) regarding availability measures improves. Use of simulations, which are partly used today already, is foreseen to increase in order to be able to design/re-design in an optimal manner during the development and operational phases of the FP lifecycle. Thus, serviceability/maintainability and reliability should necessarily be considered during the "design for diagnostics" or "design for availability", as should other emerging relevant DFx-concepts.

8.3.2 Software

The software constituent of FP is growing and is expected to continue to grow rapidly. Therefore, the software is becoming increasingly important as a lot of the functionality resides in it, and future upgrades and addition of new functionality are preferably made via additional software rather than by changing or making additions to the hardware. The following input from the respondents on the software is summarized below and further analyzed at the end of this section:

- A control system, a number of sensors, and data extractors are used to extract information from the system and store the information centrally. In this manner it is possible to deduct reference values for health indexes.
- Some customers want the data storage/analytics functionality locally, whereas other customers consider it no problem to have a centralized or cloud solution. Reasons for local software are security requirements, which vary between industries, or lack of communications/network coverage/bandwidth. One current issue with the data storage is the cost associated with large amounts of data.
- The monitoring of processes will increase and, for the analytics of some parts, data from actual measurements will be used to compare with theoretical models in order to ensure that the model and production process are aligned. Thus, there is and will be continuous learning in between the models and production processes. Even larger analyses/simulations using saved/historical data are planned to be able to find additional systematic flaws/issues, pareto errors and errors based on wrong assumptions/input. Thus, the software and analytics supports the decision making regarding availability and what measures to take. An internal cloud is used for the calculations, which provides redundancy and high availability regarding the software constituent.
- The equipment is in continuous contact with the central service system and potential problems or errors and error codes are stored in the service system. The central service system quickly relays response or work orders, often automatically, to a service/repair technician (via his/her mobile phone). Commonly, this enables the service/repair technician to be on site within an hour or two. In the future, if the number of equipment parts of functional offers sold is scaled up, a

centrally made analysis and prioritization of response/work orders may potentially become necessary.
- There is a communications connection with the machine and measurements of what the machine is used for and how much it is used are taken. This renders indications on what is worn and torn and what components, oils or parts need to be exchanged and serviced, etc.
- There have not yet been any security breaches. Satellite-links are used for internet access in some locations, which is expensive but works well. The communications set-up depends on the location and what is available, and commonly telecom operators' mobile networks (or other types of available networks) are used. Thus, there is a dependency on the telecom operators and in order to make the agreements with the telecom operators stronger, the agreements are negotiated together with our customers.
- There will be both on-board/local and centralized software to monitor and diagnose equipment. The analytic tools will reside in the (internal) clouds, and the (historical) data saved is combined with data from other sources as well, such as service records, environmental conditions, etc. A framework like this supports the back office when making, for instance, cross-searches for problems/issues and finding root causes or problem patterns.
- The software will be further modularized to enhance change management (i.e., patches and upgrades).

There is an overlap of availability measurement in the hardware (as well as the service-support system) as they are integrated in order to deliver a function to customers. Thus, some of the monitoring and analysis-related results go hand in hand with the ones from the previous hardware sub-section.

An important issue raised here is where to store and process the data generated by FP. Another issue is the connectivity, whether or not the FP is connected, what bandwidth is available and the security set up. If there is not adequate security, most processing, analytics and storage needs to be conducted locally on site—which may cause costly installations of software and servers, which in contrast to central or cloud-based storage, processing and analytics will likely require a lot more effort in terms of set-up and maintenance and cost more.

Proposed is to use data from multiple sources, process them, and combine the data in analytic methods in order to predict problems and find root causes of the problems. However, this requires a deeper understanding and knowledge of availability measures, rather than merely acting upon the symptoms and providing service or support measures without further consideration.

Automating the response to a certain level, e.g., when a known problem occurs and a service engineer needs to be dispatched, will speed up the response process (and improve the availability), save cost (less administrative staff needed) and also make the customer more satisfied.

Finally, the possibility to pin-point the need for service or replacements, etc. will change the service-support process into a needs-driven and predictive process rather than a reactive one. Having data, or facts, to use for planning of service and support actions enables optimization of planning for service engineers, procurement of spares and consumables, and it is good for the environment, as efficient route planning is possible to achieve when acting proactively instead of reactively (and when the problems or disruptions have already occurred).

8.3.3 Service-Support System

The service-support system is a large constituent which involves people, set-up of service-support organizations, spare part and consumables, knowledge management, etc., and is thus costly and needs to be optimized in order to provide the expected value to both the FP provider and customers. The requirements on availability have a large impact on how the service-support system should be planned for and set up, as large geographical areas pose tough challenges on asset management and response times (as a lot of transportation may be necessary). The following input from the respondents on the service-support system is summarized below and further analyzed at the end of this section:

- The service providers get information about (predicted) need for maintenance, changes, repairs or replacements, and contact the customers to set up a further investigation. Based on historical data and the reports from the analytic software, the service/support staff can discuss with customers and quickly pin-point most issues. This cooperation with customers, and to be able to act proactively, prior to a breakdown, is very important. Since there is a common objective, i.e., to keep the function going at an optimal level of availability versus cost, the relationship with customers is rather a partnership than a provider-customer relationship.
- Commonly, there is a service interval at x numbers of hours in operation. The further advanced monitoring and service-support systems allow for additional measures when there are signs for a needed service, etc. When contacting the customer in between regular service calls, a clear and detailed outline on what the necessary actions is required. If one is able to do this and can provide feedback on the system status after the provision of any extra services, etc., it builds trust and deepens the partnership/customer relation.
- When moving towards functions, the service jobs become additionally advanced and require thorough preparation and analytics prior to conducting the service.
- Equipment logs and stored data from operations and service/maintenance are used and combined with other relevant data in big data analytics in order to find deviations and causes of issues/problems. The data used are stored in the back-office systems, event logs (where some operational data can be retrieved) and databases. Further, other relevant data used are, for instance, GPS- and other geographical data. Further incorporated are: service data, operator-behaviour-related data (where appropriate), and results from other analyses.

- There are specialists analysing the data gathered, and special reports as well as standard KPIs are made.
- The service-support system is managed by humans, and the machines and their nominal service intervals are quite well known. This allows adequate planning for personnel demand and necessary service-support measures. In some locations, there is a need to plan for personnel exchange if there are inadequate personnel stationed in these locations. Further, if the rates of spare parts consumption per hour are known, it is quite easy to plan to have spare parts/components/oils, etc., where they are needed.
- The additions to machines and required fuel access at various locations involve dependencies on suppliers. Thus, there is a need to consider this prior to offering such extras to customers, depending on location and access to suppliers.
- There is a need to speed up the error/issue finding process by being able to use the operator, who should be able to help out and start diagnosing, before a service/support technician is dispatched.
- It will be investigated whether it is possible to use more wireless connections (i.e., mobile networks) to our equipment instead of relying on mainly wired networks, which may require additional wiring and setting up access points, routers and firewalls, etc. Using, for instance, mobile networks will make the set-up easier and faster. Further, it is necessary to get more proactive when collecting the data from equipment by better visualizing the analysis of what has been measured. The service/support data are stored centrally and analyzed centrally. Service/repair technicians can access the central service system via the internet using their laptop/tablet, and directly report which actions, measures, parts replacements, error codes, etc. have occurred. Currently, annual training on how to do the reporting is given—in order to maintain the quality and traceability. Further, there are many codes for reporting what has occurred, which aids reporting and also gives a better granularity when following up and analysing.
- All production information and outcomes of analytics will be stored in the PDM-system, in order to be able to quickly and accurately provide feedback to various production process steps, prioritize, maintain order of all documents/information/outcomes, and to make improvements to the simulation software on a long-term basis. Further, efforts are made to transfer know-how and train employees in various positions to equip them to solve problems/issues and errors faster and in a sustainable manner. There is a need to base solutions to problems on facts, rather than on assumptions, as it is then hard to find root causes.

It is clear that fact-based decision-making regarding the service-support system is wanted and necessary due to the implicit complexity and costs. To get good support for the decisions, monitoring and service-related data from (multiple) relevant sources should be compiled and combined in order to provide well-grounded facts prior to decision-making. This will enable identification of deviations and problems, as well as compilation of relevant key performance indicators (KPIs) for management level decisions. Further, simulations are needed to set up the service-support system based on the customer base, the needs, distances and the available infrastructure and staff.

Speeding up the error finding and the subsequent service-support process is of great interest if the availability level is high and disruptions are costly. Using diagnostic tools and involving the operator to a larger extent is a good idea, but that requires additional training of the operators in analytical skills and a deeper understanding of the FP.

A key to improving the knowledge of availability is to improve the quality of the data collected from various sources, but in particular from manual inputs. Training on how to input, for instance, service-related data and on spare parts will likely pay off, and it should also be possible to have a number of possible characteristics/properties to select from in the user interface as well as free-text input where nothing else is applicable. Good-quality data input will provide a basis for well-founded decisions and necessary changes.

8.3.4 Management of Operation

The management of operation constituent should manage and coordinate the hardware, software and service-support system constituents in a sustainable manner on a long-term basis. Thus, the long-term aspects must be considered, foreseen (as much as possible) and planned for. To support the considerations and planning, it is possible to use simulations and optimization tools as well as having a developed process to gather relevant intelligence from the surrounding world. The following input from the respondents on the management of operation is summarized below and further analyzed at the end of this section:

- The ability to contact the customer between regular service windows or intervals, based on facts/indications, is an important part of availability management. As a provider, it is also necessary to learn more about the customers' applications and processes in order to be able to fine tune the offer based on the knowledge gained. This would be very suitable for the OEM partners too, who may use their own software for information storage and analytics, and provide advice on, for instance, improved dimensioning and operational parameters to tune.
- Facts, based on data collected from operations, are necessary to support decisions. Guesses or assumptions should be avoided.
- After the initial installation, once operations have settled, as a provider it is essential to return to the customer and, based on data, propose potential re-fits, re-builds or re-configurations, in order to optimize the operations and energy consumption. Re-configurations may also be necessary if the customer wants to change the load or decrease/increase the production level. Both the provider and

the customer benefit from this, and these actions create a bond with the provider in fierce competition. Further, the provider-customer relation is strengthened.

- There is a dependency on the internet and the up-time of the connectivity, as there are more than a thousand pieces of equipment connected. However, the data is buffered to a certain extent if connectivity is interrupted. Usually, local providers of connectivity that can deliver the high standard required are sought. As part of the provider's engagement, access points, routers and firewalls must work at a high level of availability as well—and the communications must be of high quality.
- There is a dependency on other parties to maintain the agreed-upon level of availability, such as utilities/electricity providers (necessary to run the function) and telecommunications providers (needed for data communications and predictive maintenance), etc. Some measures to address such problems can be taken, e.g., setting up generators and enabling buffering of outgoing data to a certain extent.
- Of further interest is to assess environmental impact (i.e., lifecycle assessment analytics), particularly when the function is part of a larger system or production environment. Both the footprint and handprint are of interest, as new laws and regulations pertaining to the environmental impact are expected in the future.
- It is important that the partners, who provide parts, ensure that their parts correspond to the requirements for quality and reliability. Dependable partners are essential.
- Usage data from on-line and off-line monitoring are used and will increase. Usage data will be combined with data from other sources (for instance, service records) to become better at correlating service measures to actual use and recorded error codes. This type of analysis will enable learning if the problems/issues arise on other equipment and at other customers as well—or if the problems/issues are only relevant for the actual context investigated.
- Posed is a requirement that the equipment must last a certain number of cycles. Thus, the failure of certain parts, components or sub-systems must be tracked before expected life is through. This is important, depending on which business model is used, if the provider has the responsibility for availability. Through monitoring it can be learnt if the problem/issue is related to parts, components, sub-systems or user/operator behaviour.
- It is important to find deviations over time regarding key components, such as drive lines and gear boxes, related to e.g., inner friction—which is hard to reproduce in lab environments. It is also necessary to establish the factors and circumstances that can be regarded as fairly constant under various external conditions. This is required to find root causes, and further to characterize the boundary conditions, external loads and management, which can be considered as part of external root causes.
- There is a need to share information, knowledge, reports, error codes, etc., with the dealers/distributors in order to become better at relating arisen error codes to the systems and sub-systems. Enhanced and exact two-way sharing enables

faster pin-pointing of problems/issues and planning of service/maintenance/repair measures.

- Partly, external parties are used for service/repair, and they must meet the same requirements on quality and response etc., as the provider. This is important for the availability level and consistency.
- Cost/income is looked at continuously and, based on that ratio, long-term plans are made. Further, potential improvements, based on what has been learnt in the past, are always considered prior to new deals.
- Training and competency development, related, for instance, to administration and analytic tools, renders good effect. Potential analytics also are needed to determine whether service-support measures should be conducted a little earlier in order to achieve less wear/tear and breakdowns, and what effect that gives on a long-term basis.
- Cooperation and work are conducted together with the suppliers and customers regarding new requirements, expected changes in laws, regulations and taxation, etc., in order to allow time to manage and coordinate required changes.
- The customer needs to get timely information as to when it is time for service, replacement of parts, etc., in order to avoid emergency solutions. Maintenance strategies whereby the customers, to a certain extent, can perform the daily service (if they want to), while the provider takes care of larger maintenance/service efforts as well as replacement and production of new (expensive custom made) parts are also considered.
- New laws and regulations regarding, for instance, the handprint, e.g. diesel engine exhaust, will make the control and monitoring systems within our functions used in mobile applications/processes additionally advanced. A potential idea is to run optimization algorithms often, or on a regular basis, to fine tune operational parameters. However, that may necessitate changing operational parameters during operation.
- Greater attention to availability and redundancy during the design phase of the function would be beneficial. This also applies to the design of the customer production environment and processes.
- It is essential to base decisions on facts, not on assumptions and guesses. It will be necessary to provide more feedback to the design/development team on how various standards work out and the result of the types of solutions selected and their actual outcomes (i.e., if the solutions work out well or not prior to future customer interactions and business offers). Also necessary is improvement of the simulation models based on the feedback, and the use of simulations should be encouraged to an even greater extent. In this context, providing a function enables acquisition of information/data from the function and parts of the production process and use thereof for improvements. In addition, long-term business is good for both the provider and customer and will hopefully make the customer even more satisfied.

It is obvious that the collection of data and facts, which may be combined with estimates on future developments, will be used to improve knowledge of availability, support decision-making, and be used in simulations and optimizations during development and later operations. Optimizations regarding efficiency, energy consumption and application specifics should be conducted on a regular basis (preferably supported by simulations).

Many FP will have external dependencies on, for instance, spare parts, consumables, service engineers, telecom and connectivity providers, and the risk related to this must be communicated and adequately mitigated. However, it is not possible to mitigate all risks that are deemed as necessary to mitigate (although mitigation is always preferred) due to costs and parameters that are hard to foresee or control.

Design for availability and redundancy is needed, as it may be difficult and expensive to remedy these issues later, and both the FP and the operational environment/processes should be factored in.

The sustainability and environmental impact of FP are becoming increasingly important, and the FP footprint and handprint should be optimized during the development and operational phases.

To use FP successfully, it is necessary to allow time and flexibility to make adequate considerations prior to making changes, instead of reacting to events, so as to make changes that are sound on a long-term basis. Thus, the use of data and facts to support decisions and be able to predict and act proactively, instead of reactively, is and will become even more crucial in the future.

8.3.5 Additional Reflections Made

The respondents made a number of reflections during and after the interviews on how to improve the knowledge of FP availability and related matters. The following is a summary of these reflections which complements the results previously presented in this section:

- There is a need to become better at detecting wear and tear (which is a slow process) and be able to simulate that process in order to understand where and when it happens during the processes. Further, in the future it is of interest to add functionality (i.e., mainly via software) to the FP—which may add risks and impact availability. However, the customer requirements and the pressure to minimize cost will make this happen. Thus, it is necessary to improve the risk awareness and management further.
- Various approaches to where to keep functionality, mainly related to software, were outlined, and the decisions to favour dealing with this locally, centrally, in the cloud or by combinations of these, were based on factors such as: connectivity, security, wish to only send relevant data from the FP (unless more data are needed and full data access is wanted), regulations, laws and of course customer requirements. There are benefits with all approaches, but also various cost and effort levels required to set up and maintain the different options.

- To be able to achieve all the above, a number of skills and competencies are needed. Thus, to maintain and improve the availability successfully, as well as the knowledge of how to improve availability, will require multi-disciplinary teams.
- To increase the automation of the service-support process and the management of incidents or signs of issues, the monitoring and analytics systems need to be trained on normal situations as well as deviations. This requires advanced analytic methods and high-quality monitoring and service-related data, which is hard to achieve but necessary. In particular, getting high-quality data from external service providers, which may use their own reporting systems, is hard. As this also affects the change and configuration management of FP, it is necessary for a successful long-term operation.
- In addition, to further speed up the process of diagnosing errors or issues, the FP operators need to be trained to be able to classify errors/issues and inspect the FP. This can save a lot of time and improve the speed and accuracy of the measures invoked, particularly if spare parts are needed and there is a long distance to the customer from the service-support site.

8.3.6 Analysis—Results Seen Through the Lenses of Data and Analytic Methods

The results point to a clear trend towards increased sensoring (as well as other ways to retrieve data) and monitoring of FP and combining the collected data with data from other relevant sources in order to improve the availability, development and operation of FP. DFx-concepts such as design-for-monitoring/diagnostics/availability will get further attention, as will improvements in maintainability and reliability of critical parts of FP. Monitoring and service-related data will increasingly be used to learn more about availability and improve the overall FP and its business. The data collected (potentially combined with estimations on the future) will support decision-making, simulations and optimization pertaining to availability, efficiency, costs and energy consumption, etc.

A choice that needs to be made is where to have functionality and "intelligence" (based mainly on software) that support: the data collection, (temporary) storage, pre-processing, combining data from various relevant sources, processing, long-term data storage, analytics/decision-making and the subsequent invocation of actions. The software can be situated locally/on-board, centrally, in the cloud, or with a combination of these. There are advantages and disadvantages with all these options, and factors such as connectivity and security have a large impact on the decision as to where to have the functionality and "intelligence".

Further, there is a need to speed up the service-support process and related measures, and there are indications that parts of this process will be automated, so that decisions and actions can be invoked. However, this requires a clear understanding of what must be fixed—and thus the quality of the collected data is a must. It is necessary to ensure that the data collected via sensors, and in particular via own staff as well as external service providers, are of high quality. Maintaining the quality of the collected data and analyzed information should be encouraged by top management.

An aggregation of the input from the respondents points toward the necessity of combining monitoring and service-related data with other data such as meta-data from sub-suppliers (on e.g., reliability and performance) and data from the application in which the FP is used. The monitoring data should measure the actual pain-point, but also additional data, from other measurements, sensors or reports, with direct or indirect relations to the pain-point should be used to shed further light on the matter. Further, the use of a combination of qualitative and quantitative methods to analyze the availability is proposed. The qualitative methods may use input from the actual pain-point, combine pain-point input with other measurements, or use only measurements from parts other than the pain-point. The quantitative methods may be used to find anomalies and outliers among one or more combined measured parameters. The results of the qualitative and quantitative methods can preferably be combined where suitable to predict or mitigate problems. Thus, conducting the above data and method triangulation may increase the accuracy of predicting availability problems and also prevent serious disruptions.

Depending on the data measured, the rate and amount of data generated, and where the data are generated, a number of questions arise. Where should the data be analyzed: locally on the machine, centrally or in the cloud? The decision depends on a number of factors such as: connectivity (is the FP connected only now and then, or are the data transferred manually from the FP?) and bandwidth (i.e., how much data can be transferred?), IT/information- and cyber security, where do we have scalable processing capacity, how much data do we need to store for historical analysis, etc.? Further, another question that needs to be asked is what type of timely availability monitoring and management is required in order to honour the contractual parameter(s): on-line or continuous, hourly, daily, monthly or quarterly? If the FP have slow wear and tear processes, the timeliness can be lower. However, if the wear and tear processes can cause swift and costly disruptions, or may require time-consuming and costly replacement in case of severe breakdowns, the on-line or continuous options are attractive if they can be set up. If it is not possible to set up adequate monitoring options, the FP contract parameters should be questioned and the FP perhaps sold as something else to avoid large business risks.

Where will availability measures render the best effect? Traditionally, most measures have been applied to the hardware and service-support system constituents. However, as the software constituent is expected to grow, a lot of availability measures will be necessary in order to maintain the availability level. Further, the coordination and inter-operability of the hardware, software and service-support system (and management of operation) constituents needs to be simulated and optimized.

8.4 Conclusions and Discussion

This chapter contributes to theory by improving the knowledge of availability, through use of monitoring and service-related data, for all parts of FP and in particular the software and management of operations constituents. The software constituent is expected to grow and will thus increasingly affect the availability of the software constituent and, consequently, the overall FP. Some of the theoretical contributions may also be usable for TES and PSS/IPS^2, etc.

The practical implications are based on the availability measures listed for the four FP main constituents, and the lists should be considered as a collection of ideas and possibilities for FP providers. Both actual and planned measures, to improve the knowledge of availability, are listed for all four FP main constituents. There is a clear need and wish to collect more relevant data and analyze it, thus the design-for-monitoring/diagnosis/availability will be of greater importance for design and re-design of FP.

The managerial implications are manifold, and one top prioritization may be to adequately manage the availability issues for FP software and not wait too long before starting to do so. Further, there are ideas for availability measures pertaining to the management of operation constituent and the need to ensure the quality of collected monitoring and service data is also highlighted. Perhaps, collection of high-quality monitoring and service should be rewarded to emphasize its importance. High-quality data is crucial to be able to make the best possible decisions based on the data. In addition, new knowledge of FP availability is required, combined with additional competencies to deal with and analyze the data, in order to create value out of it. Unless value is created, the efforts will be fruitless and only considered as a cost.

A current ongoing debate is whether to have a lot of intelligence in machines, FP or similar, or have most of the intelligence outside of the machine in order to serve the whole fleet of machines. It is an advantage to have local intelligence if a lot of data generated should be pre-processed prior to being sent to some centralized processing facility. Reasons for this may be poor connectivity, that a lot of data is just unnecessary to convey, or that the machine will operate in an isolated environment and must "be able to take care of itself" by telling the operator of maintenance/replacement needs, or invoke graceful degradation, or force shutdowns, to not break down in a serious way. Further, local intelligence can be helpful to make the machine understand where it is initially installed, or if it is moved to another context, perform some self-configuration to the environment and actual application. On the other hand, a heavily centralized data management strategy will provide possibilities to detect patterns from the data generated by multiple machines, whereas data from a single machine would not be enough to draw any general conclusions.

The emerging DFx-concept design-for-sustainability gives a broad view on how to design and develop FP in a sustainable manner. It can be discussed whether the design-for-monitoring/diagnostics/maintainability/availability, etc. should be seen as a sub-DFx-concepts under the sustainability umbrella, as they all contribute to that end. If they are placed under the same umbrella, potentially together with additional ones contributing to sustainability, it could become easier to communicate that to customers, instead of having a number of various DFx-concepts to convey and explain the value of each of them.

To conclude, one question of great importance for the possibility to improve knowledge concerning availability, for FP and other related concepts, is who will own the data generated and collected, as well as who may use it how and for what? This will be an important question to sort out and be clarified in the contract and agreement between the provider (consortium) and customers.

Acknowledgements The research in this book chapter was funded by the VINNOVA Excellence Centre the Faste Laboratory and the ProcessIT Innovations R&D Centre, Luleå University of Technology, Sweden. The authors would like to thank the respondents for their interest, close co-operation and valuable feedback.

References

1. Alonso-Rasgado T, Thompson G, Elfstrom B-O (2004) The design of functional (total care) products. J Eng Des 15(6):515–540
2. Lindström J, Karlsson L, Löfstrand M, Karlberg M (2012) A development process for Functional Products: hardware, software, service support system and management of operation. Int J Prod Dev 16(3/4):284–303
3. Lindström J, Plankina D, Nilsson K, Parida V, Ylinenpää H, Karlsson L (2013) Functional Products: business model elements. In: Proceedings of 5th CIRP international conference on industrial product-service systems, 14–15th March. Bochum, Germany
4. Lindström J, Sas D, Lideskog H, Löfstrand M, Karlsson L (2015) Defining 'Functional Products' through their constituents. Int J Prod Dev 20(1):1–24
5. Lindström J, Dagman A, Karlberg M (2014) Functional Products lifecycle: governed by sustainable win-win situations. In: Proceedings of the 3rd international conference on through-life engineering services, 4–5th Nov 2014. Cranfield, UK. (Procedia CIRP)
6. Lindström J, Dagman A, Karlberg M (2015) The Functional Products technical lifecycle and its four sub-lifecycles. In: Proceedings of the 4th international conference on through-life engineering services, 3rd–4th Nov. Cranfield, UK. (Procedia CIRP, accepted for publication)
7. Parida V, Rönnberg-Sjödin D, Wincent J, Ylinenpää H (2013) Win-Win collaboration, functional product challenges and value-chain delivery: a case study approach. In: Proceedings of 2nd International Through-life Engineering Services Conference, 5–6th of November, Cranfield, UK. Procedia CIRP
8. Stahel W (1997) The functional economy: cultural and organizational change. In: Richards DJ (eds) The industrial green game: implications for environmental design and management. National Academy Press, Washington DC, USA
9. Thoben K-D, Eschenbächer J, Jagdev H (2001) Extended products: evolving traditional product concepts. In: Proceedings of 7th international conference on concurrent enterprising, 27–29th June. Bremen, Germany
10. Mont O (2001) Introducing and developing a Product-Service System (PSS) concept in Sweden. The International Institute for Industrial Environmental Economics (IIIEE), Lund University, Sweden

11. Meier H, Roy R, Seliger G (2008) Industrial product-service systems—IPS2. CIRP Ann Manuf Technol 2008:1–24
12. White AL, Sloughton M, Feng L (1999) Servicizing: the quiet transition to extended product responsibility. Tellus Inst, Boston, USA
13. Tomiyama T (2001) Service engineering to intensify service contents in product life cycles. In: Proceedings of EcoDesign 2001, 2nd International. Symposium on Environment Conscious Design and Inverse Manufacturing, 12–15th Dec. Tokyo, Japan
14. Roy R, Shaw A, Erkoyuncu JA, Redding L (2013) Through-life engineering services. Meas Control 46(6):172–175
15. Andrews JD, Moss TR (2002) Reliability and risk assessment, 2nd edn. Professional Engineering Publishing Limited, London and Bury St Edmunds, UK
16. Löfstrand M, Andrews J, Karlberg M, Karlsson L (2011) Functional Product system availability: simulation driven design and operation through coupled multi-objective optimization. Int J Prod Dev 13(2):119–131
17. Löfstrand M, Backe B, Kyösti P, Lindström J, Reed S (2012) A model for predicting and monitoring industrial system availability. Int J Prod Dev 16(2):140–157
18. Reed S, Andrews J, Löfstrand M, Karlsson L (2012) The design and modelling of maintenance during the development of functional products or product service systems. In: Proceedings of 1st international through-life engineering services conference, 5–6th Nov. Cranfield, UK
19. Reed S, Löfstrand M, Karlsson L, Andrews J (2013) Service support system modelling language for simulation driven design of functional products. In: Proceedings of 2nd international through-life engineering services conference, 5–6th of Nov. Cranfield, UK. (Procedia CIRP)
20. Reed S, Andrews J, Dunnett S, Kyösti P, Backe B, Löfstrand M, Karlsson L (2012) A modelling language for maintenance task scheduling. In: Proceedings of 11th international probabilistic safety assessment and management conference and the annual european safety and reliability conference 2012: PSAM11, ESREL 2012, 25–29th June. Helsinki, Finland
21. Kyösti P, Reed S (2015) Prediction of service support costs for functional products. Simul Model Pract Theory 59:52–70
22. Pavasson J, Ljung A-L, Karlberg M, Larsson S, Johansson S, Lundström S (2014) Challenges and opportunities with simulation-driven functional product development and operation. In: Proceedings of 3rd CIRP international conference on through-life engineering services, 4–5th Nov. Cranfield, UK. (Procedia CIRP)
23. Lindström J, Löfstrand M, Reed S, Alzghoul A (2014) Use of cloud services in Functional Products: availability implications. In: Proceedings of the 6th CIRP international conference on industrial product-service systems, 1st–2nd May. Windsor, Canada. (Procedia CIRP, 16)
24. Reim W, Parida V, Lindström J (2013) Risks for Functional Products—empirical insights from two swedish manufacturing companies. In: Proceedings of the 2nd international through-life engineering services conference. (Procedia CIRP, 11, 340–345)
25. Schwabe O, Shehab E, Erkoyuncu J (2015) Long tail uncertainty distributions in novel risk probability classification. In: Proceedings from 3rd CIRP global web conference, 3rd–5th June, 2014. (Procedia CIRP, 28, 191–196)
26. Rajagopal S, Erkoyuncu JA, Roy R (2015) Impact of software obsolescence in defence manufacturing sectors. In: Proceedings from 3rd CIRP global web conference, 3rd–5th June, 2014. (Procedia CIRP, 28, 197–201)
27. Okoh C, Roy R, Mehnen J, Redding L (2014) Overview of remaining useful life prediction techniques in through-life engineering services, product services systems and value creation. In: Proceedings of the 6th CIRP conference on industrial product-service systems. (Procedia CIRP, 16, 158–163)
28. Datta PP, Roy R (2010) Cost modelling techniques for availability type service support contracts: a literature review and empirical study. CIRP J Manuf Sci Technol 3(2):142–157
29. Reim W, Parida V, Rönnberg Sjödin D (2016) Risk management for product-service system operation. Int J Oper Prod Manag 36(6):665–686

30. Vasantha GVA, Roy R, Lelah A, Brissaud D (2012) A review of product–service systems design methodologies. J Eng Des 23(9):635–659
31. Tukker A (2013) Product services for a resource-efficient and circular economy—a review. J Clean Prod 97:76–91
32. Patton MQ (1990) Qualitative evaluation and research methods. Sage Publications, London, UK
33. Kvale S, Brinkmann S (2009) InterViews: learning the craft of qualitative research interviewing. Sage Publications, LA, USA
34. Yin RK (2003) Case study research: design and methods. Sage Publications, Thousand Oaks, CA, USA
35. Fontana A, Frey J (1994) Interviewing. In: Denzin N, Lincoln Y (eds) Handbook of qualitative research. Sage Publications, Thousand Oaks, CA, USA
36. Miles M, Huberman M (1994) An expanded sourcebook—qualitative data analysis, 2nd edn. Sage Publ, Thousand Oaks, CA, USA

Chapter 9
Remodelling of Structured Product Data for Through-Life Engineering Services

Sebastian Adolphy, Hendrik Grosser and Rainer Stark

Abstract Product data of long-living and complex systems is essential for efficient through-life engineering services (TES), since it is the basis for systematic planning, operation and documentation activities. The product structure represents the data backbone and carries geometrical, technological and system relevant information. Product Lifecycle Management (PLM) systems are used to maintain the product data and to track product changes. However, in the maintenance, repair and overhaul (MRO) sector lifecycle based documentation of changes regarding product configuration, condition and functionality is still a weak spot, especially if the MRO service provider is not the original manufacturer. In such cases MRO processes start with an exhaustive product diagnosis to identify parts that have to be maintained or to determine if spare parts are needed to guarantee product's performance. Existing 3D scanning and data processing methods have to be improved to acquire structured product data in an efficient process. This chapter presents a method for automated derivation of product structures from 3D assembly models and its application in various scenarios of through-life engineering services. The input may be a system model coming from a 3D scanning process. In order to identify spatial relations between parts a 2D contact graph is automatically created by a neighborhood analysis. Subsequently related parts are hierarchically structured into sub-assemblies by a cluster analysis. Iterative use of these two principles results in the complete product structure. Results can be exported via a XML interface for use in PLM systems.

S. Adolphy · H. Grosser · R. Stark (✉)
Division of Virtual Product Creation, Fraunhofer IPK, Berlin, Germany
e-mail: rainer.stark@ipk.fraunhofer.de

S. Adolphy
e-mail: sebastian.adolphy@ipk.fraunhofer.de

R. Stark
TU Berlin, Berlin, Germany

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_9

9.1 Introduction—Structured Product Data and Its Remodelling

The method proposed in this chapter addresses an issue which is not predominant in the domain of through-life engineering services. For a general understanding this introductory section will answer to questions to the reader:

1. What is structured product data?
2. How is product data structured?

9.1.1 What Is Structured Product Data?

During the product creation process using computer aided design (CAD) and engineering (CAE) methods various data is created in order to build models of the emerging product. These models evolve in an evolutionary process from merely text based descriptions of requirements up to highly functional virtual models of the system. Advanced design methods are not only able to produce geometrical models of the system, but to build "digital twins", which mirror the behaviour of the system into the virtual world before it even came to existence in reality. Virtual product models require the incorporation of a high number of different data elements [1]. The sum of this entire product related data is here referred to as "product data". The product does not remain in a static after its definition in design. The production process may result in variations, during its use the product may be altered and service activities may include modifications. In an ideal situation (Fig. 9.1) this evolution of the product throughout its lifecycle is mirrored

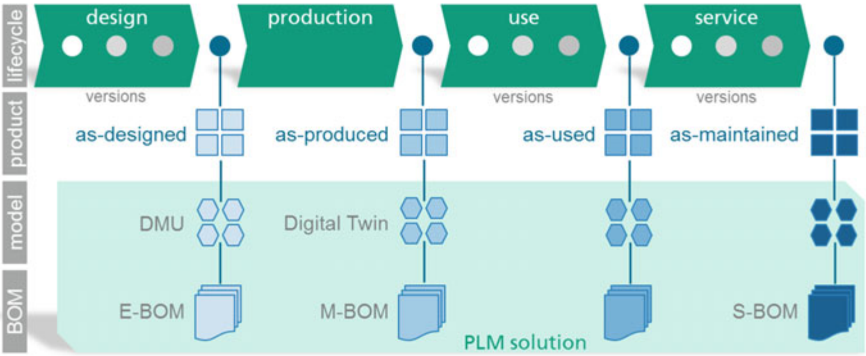


Fig. 9.1 Ideal situation for availability of product data throughout the product lifecycle

in the corresponding product data, foremost the model and bill of material (BOM) representing the current state of the product. Of course this ideal situation is not always met in industrial practice.

9.1.2 How Is Product Data Structured?

In order to store, manage and retrieve the product data it is commonly structured in relational databases. The product data structure—commonly shortened to "product structure"—can be, but does not have to be similar to the building structure of assemblies, subassemblies and parts of the physical system. Structuring elements can also be stages of production processes or functional units. The favored approach is dependent on the primary stakeholder of the structured product data [2]. For the initial generation of the product structure in design the two major rivalling perspectives are those of engineering and manufacturing. Whereas designers prefer assembly or functional oriented structures as this complies with their thinking about products, manufacturers favor structures mirroring the production process. As other stakeholders in later lifecycle phases—e.g. MRO and other services—have divergent requirements, the use of different product structures for the same system throughout its lifecycle is common. Workload for creation, conversion, updating and linking of these product structures cumulates in a substantial matter of expense.

Definition of product structures in PLM systems is a manual task if it cannot be retrieved from an assembly structure in a linked CAD system. Large corporations require departments of "structure managers" for product structure generation and management. They perform an extensive, highly manual process, sometimes totally separating engineers from the PLM environment:

- Single part files have to be coded following a company specific naming and structuring convention as part of an internal enterprise information standard. This is done by creation of PLM structure templates which include main structure nodes representing sub-assemblies and related parts. These templates are generic, but for each product it has to be decided which nodes are used.
- CAD models in order to be processed have to be saved in a proprietary file format with a specific title and ID according to the structure template.
- Afterwards all CAD files are imported into the PLM system.
- For high-performance visualization additional files such as JT formats have to be created from the proprietary files and stored with the correct data set.

This highly manual process is of course both error-prone and resources-intensive.

9.1.3 What Is Remodelling?

"Remodelling" in this context is meant to describe the retroactive recreation of a virtual product model from the physical system.

The term has been chosen in order to enable differentiation from methods of reverse engineering, with their negative connotation due to links to intellectual property violations. It is not purpose of remodelling to recreate a physical clone of the product. The virtual product model created by remodelling is used to enable activities supporting the original product.

Today comprehensive knowledge on the respective product is needed to manually model a proper product structure in order to provide structured product data for through-life engineering services [3]. Automatic identification of product structures for 3D product models currently requires the availability of assembly structure as defined in the authoring CAD system. For long-life systems and third party service contractors this knowledge and data is usually not available.

Most research on reverse engineering is focused on recognition of surfaces rather than on entire systems and their structures. Approaches are based on basic shapes like spheres, cylinders, cuboids [4] or other features and constraints [5]. Segmentation methods are used to detect edges and surfaces [6, 7], not assemblies and their parts. In the construction domain the retrieval of single object information is pursued in order to create unstructured component lists [8, 9] by Building Information Modelling (BIM) applications used for reconstruction of large scale infrastructure. Despite attempts for semi-automated features for reverse engineering the high effort still poses an barrier for wide spread application [10–12].

Primal result of 3D scanning methods is a cloud of data points which include x-, y-, z-information as well as a normal direction. Such a point cloud is then transformed to a mesh or polygon surface. The meshing commonly requires post-processing as shiny or concave surfaces result in incomplete meshes and artefacts. For the further processing and modification in CAD systems surface reconstruction is necessary, resulting in Non-Uniform Rational B-Spline (NURBS) surfaces.

The remodelling process to create assembly models consisting of separate parts contains the following steps:

1. 3D scan of the complete assembly,
2. 3D scans of each single disassembled part,
3. 3D polygon models of the parts are referenced to the initial assembly scan (requires reference markers on the object's surface).

For the additional creation of CAD assembly models the single part's 3D polygon models have to be converted to surface models. These CAD parts are manually assembled in a CAD tool. More details on an enhanced process for automated reverse engineering of assembly models are described by [13].

9.2 Motivation—Need and Availability of Structured Product Data in Through-Life Engineering Services

After the introduction of structured product data as the subject matter, this section will answer the two following questions in order to explain its relevance for through-life engineering services:

1. Why do through-life engineering services need structured product data?
2. Why does the product data have to be remodelled?

9.2.1 Why Do Through-Life Engineering Services Need Structured Product Data?

Through-life engineering services benefit largely if structured product data is available and the product data structure can be employed for the structured storage of data created within these activities:

- **Inspection and planning**
  Assessing degradation of a component in service using non-destructive evaluation techniques and automating the assessment process are two major trends in maintenance [14]. An as-designed model of the system can be consulted as reference for analysis of changes of the system throughout its lifecycle. This deviation analyses can even be automated, if an as-used model is created by methods of 3D scanning. In case the as-designed model is not at hand, it can either be automatically searched in a database by the created as-used model, or the latter can be used to recreate the as-designed model, for example by statistical analysis of a great number of identical used parts, e.g. turbine blades.
- **Repair, overhaul and replacement of parts**
  van Houten et al. [15] identified the product data model to support model based maintenance planning. The aforementioned deviation analysis will allow defining repair and overhaul measures. In case the damaged part has to be replaced, reproduction of spare parts not in stock and for which no as-designed model exists will be enabled by the created models. In common practice the parts are “re-designed” manually, which is time consuming and error prone.
- **Monitoring and documentation**
  Throughout its lifecycle further product related data is created, describing either the status of the product itself or accompanying processes involving the product. Such lifecycle information can be quality data from the production process, changes in hard- or software configuration or reports on through-life engineering services. Managing lifecycle data is essential for an integrated maintenance planning capability [16]. Storage and access to such data is supported by linking it to product structures.

9.2.2 Why Does the Product Data Have to Be Remodelled?

In today's industrial practice the structured product data as created during the design process is often not transferred and employed in later stages of the product lifecycle. As well lifecycle based documentation of changes regarding product configuration, condition and functionality is still an unsolved problem for holistic product lifecycle management approaches [17]. Common barriers are organizational structures within or between companies, the use of segregated IT systems and data incompatibility, particularly problematic for long-life systems.

In traditional hardware-sales-driven business models the designing and producing companies are not closely involved in the operation and service of their products. Operators and third party service provider on the other hand do not have access to the original product data. Even if the product lifecycle lies within the hands of one company, isolated departmental structures can have similar effects.

The different roles allocated to the product lifecycle in Fig. 9.2 are not mutually exclusive inherited by separate players, but also one role may be split up among several players. Numerous constellations can be found, e.g.:

- Design and manufacturing may be distributed among several tiers of suppliers.
- Nevertheless design and manufacturing are often controlled by one OEM, with more or less portions of both being outsourced.
- This OEM may even stay owner and operator of the product, while the user is a customer.
- Many OEMs also act as service providers, for maintenance or in context of product service systems.
- But then again the user may also take care of operation and maintenance by himself, or employ a third party service provider.

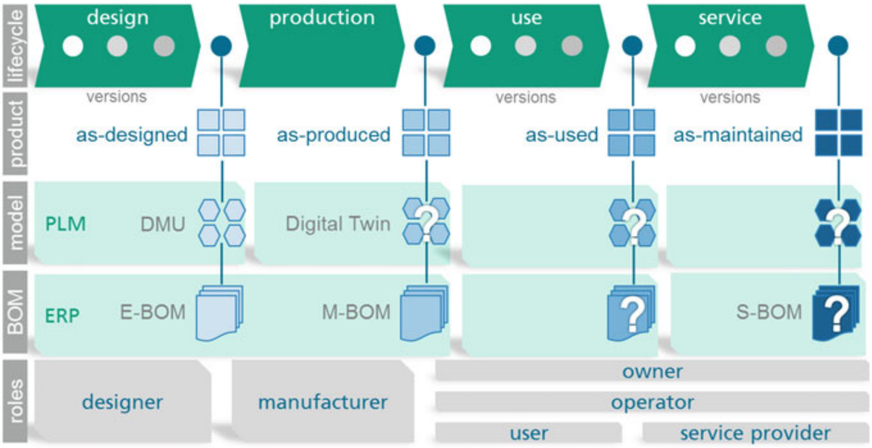


Fig. 9.2 Roles and access to product data throughout the product lifecycle

Depending on the constellation of players and their business relations, reality of product data availability may vary vastly from the ideal situation.

Aside from the organizational barriers the use of isolated or sparsely related IT systems is the major reason for discontinuity in data flow throughout the product lifecycle. PLM-Systems thrive to cover the entire product lifecycle and have over the last years incorporated many additional features supporting activities traditionally solved by independent software tools. Thoroughly permeation of the product lifecycle nevertheless is still to come. Companies fear the vulnerability of monolithic IT solutions and the dependency on their vendors. Existing landscapes of diverse IT systems are robust and any changes are risky.

The situation of data availability in many cases is unsatisfactory, as there are many TES actives which would largely benefit from the availability of structured product data. Efficient data reconstruction or retrieval strategies are needed [13].

9.3 Method—Remodelling of Structured Product Data

The developed method for automated generation of product structures is a two-step procedure.

Step 1: Identification of spatial relations between parts by neighborhood analysis using contact graphs

A graph G is an abstract datatype which represents a structure. It consists of a finite set of vertices (or nodes) V and edges E. Additionally, it is possible to assign attributes η to edges and v to vertices given a node label alphabet L_v and an edge label alphabet L_e [18].

$$\begin{gathered} G = \{\{V, E, \mu, \eta\} | v, e \in \mathbb{R} \\ \eta : v \to L_v; \mu : e \to L_e \end{gathered} \tag{9.1}$$

Applied to product structures, vertices represent parts of the product while an edge represents a direct spatial contact between two parts. The relations between the parts are saved in an adjacency matrix or list. In a basic form the attributes of a vertex only consist of its name and an ID for internal processing. The undirected edges of the graph own no attributes; this means the only information contained is the existence of a connection between two parts.

The contact graph is created by analysing the neighborhood of the parts of a product in its 3D model. Geometric 3D models can be represented in various formats. It is important to consider the possible input formats before deciding about a transformation method. This method uses tessellated surface models as input data,

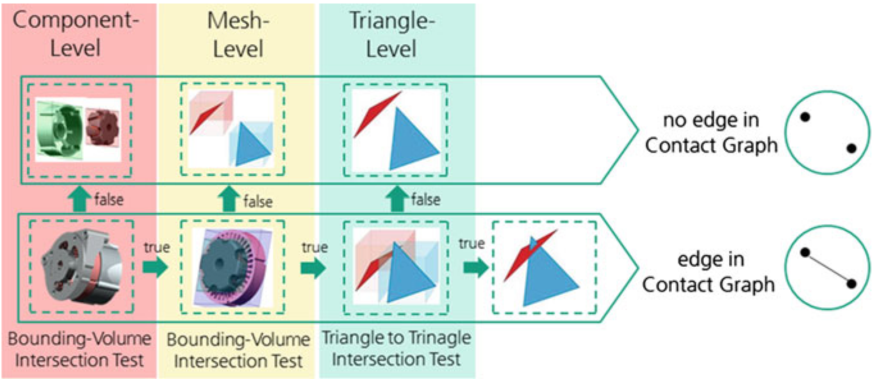


Fig. 9.3 Three stage intersection test method for neighborhood analysis

which requires the complete transformation of the point cloud generated via scanning into a global coordinate system. Goal of the neighborhood analysis is to identify all possible spatial contacts between the product parts. In order to reduce computing time by avoiding superfluous comparisons of triangles the method uses a three-stage neighborhood analysis.

The first two stages on component and mesh level as shown in Fig. 9.3 are based on the highly simplified axis alignment comparison of bounding boxes, which reduces the computation time significantly. The last stage consists of tri-tri Möller intersection tests of two triangles at a time [19]. The procedure starts with assigning bounding volumes to each part of the product. If they possess no intersection there is no contact edge between these parts and the analysis can turn to other parts. If they intersect, the level of detail is increased for the analysis of this segment. Now the bounding volume intersection test is repeated on the mesh level. This means, bounding volumes are assigned to every triangle of the two considered parts and mutually tested for intersection. If each intersection test is negative there is once again no edge between these parts. If there is an intersection a detailed tri-tri Möller algorithm is applied to these two triangles for verification of the contact.

Step 2: Hierarchical structuring of related parts by cluster analysis

In the second step the derived contact graph is subject to a cluster analysis. By applying a cluster algorithm to the graph it is possible to divide it in several subsections. In order to build a product structure with more than one level it is necessary to apply a cascading sequence of clustering steps in a top-down process (Fig. 9.4).

Graph theory provides different algorithm for clustering. Based on the assumption that assemblies as the structural groups of a product structure are well-separated modules with sparse connection between each other, the Girvan-Newman algorithm was selected. This algorithm has been developed for

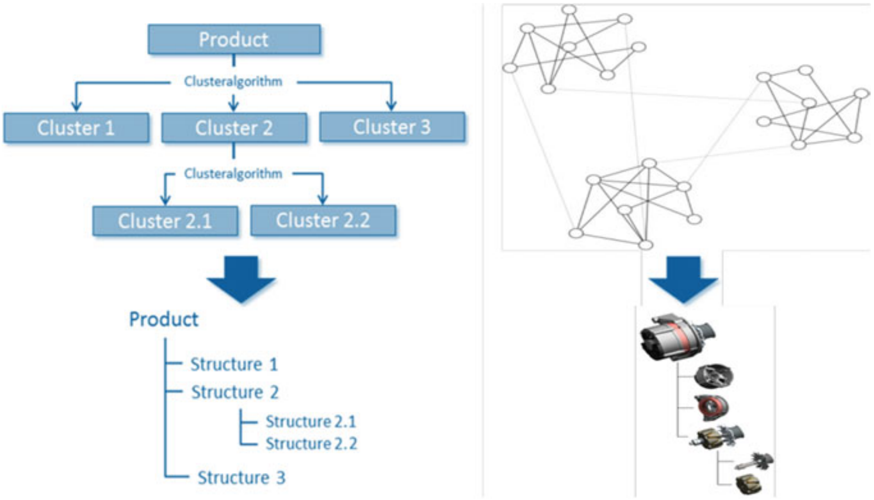


Fig. 9.4 Top-down process for product structuring by cluster analysis

identification of community structures in social and biological networks. It is based on the “edge betweenness” which is calculated for every edge of the created graph.

Edge Betweenness is a measure of an edge’s centrality in a network. It is equal to the number of shortest paths from all vertices to all others that pass through that edge. A high value of edge betweenness is an indicator for connecting edges between two communities.

The edge betweenness is calculated by the sum of shortest paths between pairs of vertices that run through the edge. The communities are formed by progressively deleting the edges with the highest betweenness value and recalculating the betweenness values. The stop mechanism can be set by a betweenness-threshold or a minimum number of communities [20]. This cluster analysis is resulting in a single graph only. To build a product structure with hierarchical levels it is indispensable to repeat the cluster analysis on every node of the future product structure. A manual check for correctness to enable potentially needed modifications is recommended after each cluster analysis. For example very small gaps between parts may mistakenly result in a contact area if the product model was created by 3D scannig.

The top down procedure of structuring the product hierarchically (Fig. 9.5) starts with the root node of a product structure which represents the entire system. Afterwards, those clusters are selected which require further subdivision.

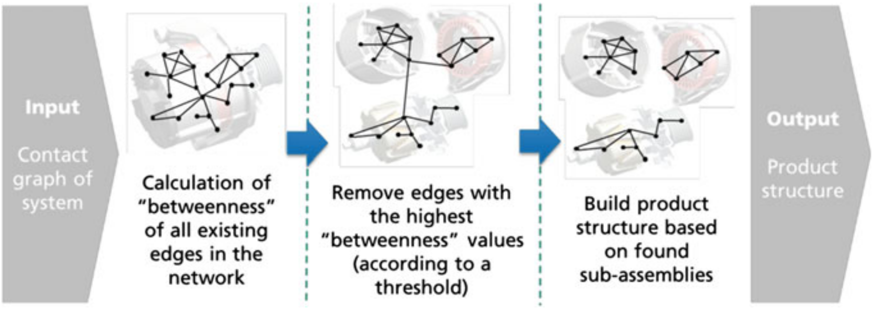


Fig. 9.5 Hierarchical structuring of related parts by cluster analysis

9.4 Application—Implementation and Use Cases

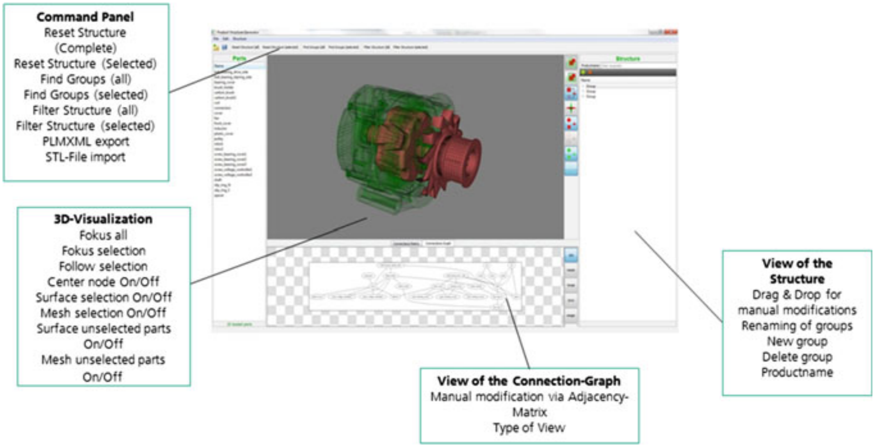
For its application the method has been implemented in a software prototype, which consists of five functional elements:

1. STL Importer: Parsing of single part STL files.
2. Neighborhood Analyser: Creation of contact graph from imported STL files.
3. GN-Cluster Analyser: Division of selected graphs into subgraphs.
4. Graphical User Interface:
    - 3D visualization for manual group selection and corrective actions,
    - product structure display with highlighted selected structure knots via 3D visualization,
    - drag and drop feature for manual manipulation of product structures.
5. PLMXML Exporter: Enables combined and comfortable import of the derived product structure and the STL files in a PDM system (Fig. 9.6)

As outlined in the motivational section, there are various constellations of both required and available product data. If there is no product data available at all, the processes of recreating those are much more elaborate, since the as-designed model has to be derived from the current system status. If service providers can access at least as-designed models of the product, this will help to accelerate proceedings substantially.

9.4.1 Use Case 1: Remodelling of As-Used Systems

The major use case for the developed method is the creation of virtual as-used models of the system as input for various through-life engineering services. The 3D scanning of defective systems enables detection of deviations between as-used and as-design conditions. If as-designed models are not available they can be recreated



Command Panel
Reset Structure (Complete)
Reset Structure (Selected)
Find Groups (all)
Find Groups (selected)
Filter Structure (all)
Filter Structure (selected)
PLMXML export
STL-File import

3D-Visualization
Fokus all
Fokus selection
Follow selection
Center node On/Off
Surface selection On/Off
Mesh selection On/Off
Surface unselected parts On/Off
Mesh unselected parts On/Off

View of the Connection-Graph
Manual modification via Adjacency-Matrix
Type of View

View of the Structure
Drag & Drop for manual modifications
Renaming of groups
New group
Delete group
Productname

Fig. 9.6 GUI of the software prototype

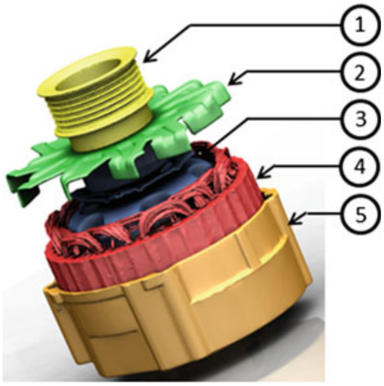


Fig. 9.7 3D model of a car alternator generated from 3D scan data (*1*. Pulley, *2*. Fan, *3*. Rotor, *4*. Inductor, *5*. Housing)

based on the digitized as-used system. When relevant deviations are found the digitized as-used system model can be employed in the following for the definition of repair measures or spare part production. As defective systems are rarely isolated components, structuring of the digitized system is necessary. The ability to generate structured models by 3D scanning systems without prior disassembly is highly favorable in terms of downtime reduction [13].

Step 1—Neighborhood analysis using contact graphs—is demonstrated by application of the method to a 3D scanned car alternator (Fig. 9.7).

The parts have been optically 3D scanned separately and automatically matched in a scan of the assembly by a best-fit algorithm. The prepared STL files of the referenced parts are imported and analysed in the software prototype, resulting in adjacency matrix (Table 9.1).

Tab.9.1 Adjacency matrix of car accelerator

	Fan	Pulley	Housing	Rotor	Inductor
Fan		x		x	
Pulley	x			x	
Housing					x
Rotor	x	x			(x)
Inductor			x		

The connections have been analysed correctly, except for the connection between rotor and inductor. The misleadingly identified connection between these parts resulted from a very small gap in the 3D model between these parts which implies high demands on the referencing process.

9.4.2 Use Case 2: Structuring of 3D CAD Models for Deviation Analysis

Structuring of given 3D CAD models is necessary when an as-used model generated by 3D scanning has to be compared with the related as-designed model. A software supported comparative analysis of product models requires identical product structures. By application of the proposed method identical structures can be generated for both the original and the scanned model.

Demonstrated on the as-designed model of the introduced car alternator—imported as 25 separate tessellated STL files—the part's neighborhood could be detected without errors in the first step. When applied to the root node in the second step, the clustering algorithm automatically allocated the first level of the product structure in three different groups. The first and second clusters are defined by the static parts grouped respectively to the two-parted-housing. The third group is characterized by the rotating parts of the alternator, which is subjected to a repetition of the procedure in order to validate the generation of a multi-level product structure. The algorithm separated the coil and its holders from the rest of the group. The entire structure as shown in Fig. 9.8 was created without manual post-editing.

9.4.3 Use Case 3: Restructuring of Product Data for Deviation Analysis

A special variant of the aforementioned use case is a situation with an existing model coming with a given product structure in a PLM system. Unless the given structure has been generated analogue to the clustering algorithm in the first place a restructuring of the product data is prerequisite for the model-based failure

Fig. 9.8 Structured as-designed model of car alternator

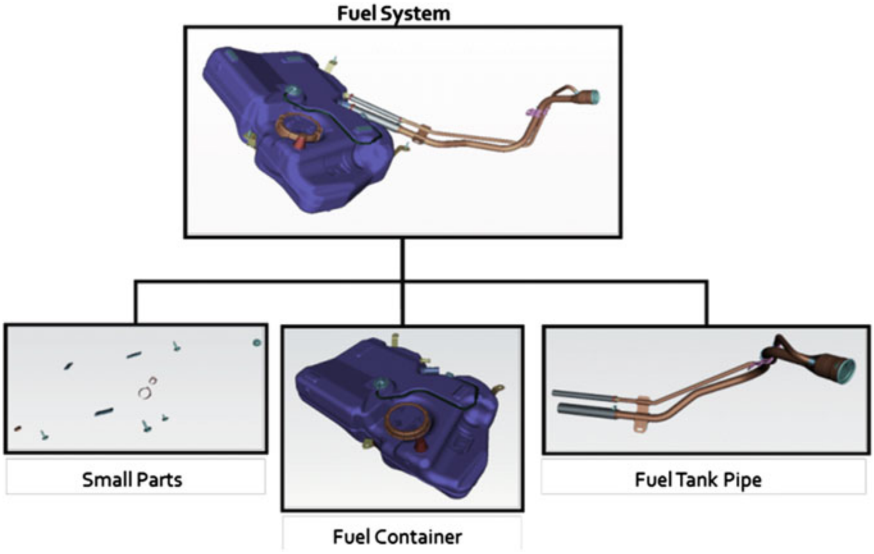


Fig. 9.9 Target groups for clustering of a fuel system

detection in MRO processes. This use case is demonstrated by the clustering of a fuel system. The as-designed model and its product structure were provided by an automotive manufacturer. The model consists of 37 parts which should be clustered in three groups as shown in Fig. 9.9.

As the algorithm is not able to identify the structural unconnected small parts (joining elements), the parts of this group were excluded from the analysis. Thus, only the fuel tank assembly and the tank pipe assembly had to be found.

The algorithm created three clusters. The two assemblies were clustered correctly, except for one part each, which has been assigned to the wrong assembly respectively. The third cluster contained only a single part, which should have been included in the tank pipe assembly.

More details of the tests run on the software prototype related to the three described use cases have been described in [3].

9.5 Conclusions and Outlook

Concluding on the current status of the method and the options for further development two questions have to be answered:

1. What is left to do?
2. What are the future prospects?

9.5.1 What Is Left to Do?

9.5.1.1 Further Automation for Industrial Application

The neighborhood analysis works error-free, as long as the imported model is error-free as well. Thus, if the 3D assembly model is derived from a 3D scanning process the results have to be precise and reliable to avoid manual corrections. The multi-step process as implemented in the prototype contains some need for human interference as well. The top-down process is a possible way to create a multilevel product structure. This process relies on a manual selection for further clustering. An industrial application of the method would call for a further automation of the process in order to increase the efficiency and reduce the potential for failures in the operation.

9.5.1.2 Up-Scaling to High Complexity Systems

The method has been validated by a prototypical implementation, which was successfully applied in various use cases. These test cases have been sufficient for the evaluation but do not inherit the complexity of many technical systems subject to through-life engineering services. The large scale evaluation on highly complex system will lead to further development of the method.

9.5.2 What Are the Future Prospects?

9.5.2.1 Closing the Gaps in Product Lifecycle Management

The term Product Lifecycle Management (PLM) is meant to express the continuous support of a products lifecycle with its related processes, information and data [21]. Although the term PLM is widely used both by software vendors and their customers in industry, PLM-systems are often primarily used for team data management (TDM), restricted to the management of geometrical models within the design department. The ability to close gaps in the data flow throughout the product lifecycle by the developed method could be a chance to foster the dissemination of the PLM approach.

9.5.2.2 Digital Twins for Continuous and Smart Maintenance

Digitalization of products and their production processes is the dominant technological trend in the second decade of the 21st century. Technological developments such as Internet of things, Cyber Physical Systems, Smart Products and the Industrial Internet or Industry 4.0 are building blocks of an interconnected world of technical systems. In order to create and operate these systems digital twins are required, enabling a paradigm shift in our ability to better understand the health of a product and plan maintenance based on the availability of significantly large volume of data [14]. Therefor the role of virtual product models will increase, predictably spreading their use over the entire product lifecycle. The inevitable integration of existing long-life systems into this connected world and the application of smart maintenance methods to them will require the remodelling of structured product data.

9.5.2.3 Big Data Analytics for Structured Product Data

The developed method is not limited to the presented applications within through-life engineering services. An interesting future area of application could be a comparison of products by means of pattern matching between two contact graphs. This could enable fast query search of structural related products in big data stocks. As the algorithms employed in the method lead to reproducible results, the method could be applied to systematically create original product structures from CAD data, in order to eliminate human-induced variations in product structures. The same principle could be applied to restructure product in order to achieve comparability. An application for this approach could be a metadata independent product variant analysis, e.g. in order to reduce repetitive design of redundant components. The large scale data analysis would be enhanced by the facilitation of data exchange between prior incompatible software systems or structure conventions by provision of a platform independent template structure in the PLMXML format.

References

1. Stark R et al (2010) Competing in engineering design—the role of virtual product creation. CIRP J Manuf Sci Technol 3(3):175–184. http://www.sciencedirect.com/science/article/pii/S1755581710000854
2. Eigner M, Stelzer R (2009) Product lifecycle management: Ein Leitfaden für product development und life cycle management, 2nd edn. Springer
3. Stark R, Grosser H, Müller P (2013) Product analysis automation for digital MRO based on intelligent 3D data acquisition. CIRP Ann Manuf Technol 62(1):123–126
4. Wang J, Gu D, Yu Z, Tan C, Zhou L (2012) A framework for 3D model reconstruction in reverse engineering. Comput Ind Eng 63(4):1189–1200
5. Kang L, Li Y, Chen ZM (2012) An integrated reverse modeling approach based on reconstruction of features and constraints. AMM 215–216:639–642
6. Ma T-C et al (2012) Features detection from industrial noisy 3D CT data for reverse engineering. In: Lee R (ed) Studies in computational intelligence, software and network engineering. Springer, Berlin, pp 89–101
7. Digne J, Morel J-M, Mehdi-Souzani C, Lartigue C (2012) Mesh segmentation and model extraction. In: Hutchison D, Kanade T, Kittler J, Kleinberg JM, Mattern F, Mitchell JC, Naor M, Nierstrasz O, Pandu Rangan C, Steffen B, Sudan M, Terzopoulos D, Tygar D, Vardi MY, Weikum G, Boissonnat J-D, Chenin P, Cohen A, Gout C, Lyche T, Mazure M-L, Schumaker L (eds) Lecture notes in computer science, curves and surfaces. Springer, Berlin, pp 236–252
8. Oreni D, Brumana R, Cuca B (2012) Towards a methodology for 3D content models: the reconstruction of ancient vaults for maintenance and structural behaviour in the logic of BIM management. In: 2012 18th international conference on virtual systems and multimedia (VSMM), pp 475–482
9. Ehm M, Hesse C (2014) 3D-Laserscanning zur Erfassung von Gebäuden—Building information modeling (BIM). Bautechnik 91(4):243–250
10. Yoon S, Jung J, Heo J (2015) Practical implementation of semi-automated as-built BIM creation for complex indoor environments. Int Arch Photogramm Remote Sens Spatial Inf Sci XL-4/W5:143–146
11. Hidaka N, Michikawa T, Yabuki N, Fukuda T, Motamedi A (2015) Creating product models from point cloud of civil structures based on geometric similarity. Int Arch Photogramm Remote Sens Spatial Inf Sci XL-4/W5:137–141
12. Barazzetti L et al (2015) BIM from laser clouds and finite element analysis: combining structural analysis and geometric complexity. Int Arch Photogramm Remote Sens Spatial Inf Sci XL-5/W4:345–350
13. Grosser H, Stark R (2014) Advanced 3D scan data analysis for performant reengineering maintenance processes. In: Farinha JT, Galar D (eds) Proceedings of maintenance performance measurement and management (MPMM) conference 2014. Imprensa da Universidade de Coimbra, pp 73–80
14. Roy R, Stark R, Tracht K, Takata S, Mori M (2016) Continuous maintenance and the future—foundations and technological challenges. CIRP Ann Manuf Technol 65(2):667–688
15. van Houten FJAM, Tomiyama T, Salomons OW (1998) Product modelling for model-based maintenance. CIRP Ann Manuf Technol 47(1):123–128. http://www.sciencedirect.com/science/article/pii/S0007850607628005
16. Ni J, Jin X (2012) Decision support systems for effective maintenance operations. CIRP Ann Manuf Technol 61(1):411–414. http://www.sciencedirect.com/science/article/pii/S0007850612000674
17. Stark R, Grosser H, Beckmann-Dobrev B, Kind S (2014) Advanced technologies in life cycle engineering. Procedia CIRP 22:3–14

18. Cook DJ, Holder LB (eds) (2007) Mining graph data. Wiley-Interscience, Hoboken, NJ
19. Möller T (1997) A fast triangle-triangle intersection test. J Graph Tools 2(2):25–30
20. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci 99(12):7821–7826
21. Stark R et al (2011) Virtuelle Produktentstehung in der Automobilindustrie. Informatik Spektrum 34(1):20–28

Chapter 10
Holistic Approach for Condition Monitoring in Industrial Product-Service Systems

Eckart Uhlmann, Christian Gabriel, Abdelhakim Laghmouchi, Claudio Geisert and Niels Raue

Abstract The content of the book chapter is the development and application of a method for the cost optimized integration of a condition monitoring system for machine tools. An environment for the simulation of lifecycle costs considering the maintenance processes as well as the hardware will be described to identify an appropriate sensor concept. The approach of event based simulation allows an assessment of possible sensor concepts depending on the machine tool's performance. To realize a cost-effective condition monitoring solution using simple consumer electronics, such as Micro-Electro-Mechanical-Systems, and to provide high scalability a wireless sensor network has been developed and evaluated. It can be easily adapted to different specific applications because of decentralized data preprocessing on the sensor nodes, as well as services in the cloud. Within this network the sensors interact through a software agent system which is implemented in the machine tool and all of its subsystems. The Java Agent Development Framework will be used as a middleware. The modularization leads to a highly flexible system. Additionally, the agent system enables the interaction between the machine tool, the IPS^2 provider, and its service technicians. The book chapter includes the evaluation of the method in the field of grinding machine tools by means of a feed axis.

E. Uhlmann (✉) · A. Laghmouchi · C. Geisert
Fraunhofer Institute for Production Systems and Design Technology IPK,
Pascalstrasse 8–9, 10587 Berlin, Germany
e-mail: eckart.uhlmann@iwf.tu-berlin.de

A. Laghmouchi
e-mail: abdelhakim.laghmouchi@ipk.fraunhofer.de

E. Uhlmann · C. Gabriel · N. Raue
Institute for Machine Tools and Factory Management, TU Berlin,
Pascalstrasse 8–9, 10587, Berlin, Germany

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_10

10.1 Introduction

The perception of industrial services has changed during the past years. Nowadays, services are an integral part of the offer to fulfill the customers' demands. Therefore, services are no longer seen as an add-on to the core product. This development is due to two causes. Firstly, in the service sector increased margins can be achieved compared to the selling of products [5]. Secondly, the consistent orientation on customer needs leads to a stable customer provider relationship and thus advantages for all stakeholders involved.

Approaches for the transformation of companies from a product seller to a solution seller, where industrial services are an equal part, are subsumed under term servitization [10]. Product Service-Systems (PSS) are a particular form of servitization as the focus lies on the use of product shares and not merely on the selling of products. As a consequence, the customers benefit from a restructuring of the risks, responsibilities, and costs traditionally associated with ownership [1].

PSS for the industrial sector are called Industrial Product-Service Systems (IPS^2) and are characterized by the integrated and mutually determined planning, development, provision, and use of product and service shares including its immanent software components in Business-to-Business applications and represents a knowledge-intensive socio-technical system [5]. The individuality of IPS^2 can lead to innovative IPS^2 business models, since the IPS^2 provider guarantees specific machine tool availability or certain amounts of parts manufactured by the machine tool [5]. Therefore, the IPS^2 provider faces difficulties in predicting and controlling risks and uncertainties that usually would be the problem of the customer [8].

10.2 Objectives and Approach

IPS^2 allow the individual configuration of solutions consisting of product and service shares which is an advantage for customers towards the classical approach of buying goods. Therefore, a wide variety of external and internal influences on customer site have to be taken into account in the machine tool industry. In the case where the IPS^2 provider covers the guarantee of machine tool availability an appropriate and customer individualized concept for maintenance, repair, and overhaul (MRO) activities is crucial.

One objective of this book chapter is the development of a holistic approach to develop and simulate a customer provider relationship where a specific machine tool availability is guaranteed. Thus, the IPS^2 provider is able to quantify the effort considering the MRO costs over the entire lifecycle. In the development phase a business process model and notation (BPMN) based workflow management system is used to model and simulate the necessary activities for different MRO concepts.

These concepts differ regarding the existence of a condition monitoring for proactive maintenance or regarding the methods for spare parts supply. The cost efficiency of these concepts depends on different factors on customer site. Machine tools, which operate in a three-shift production, have different requirements concerning MRO concepts than machine tools in a one-shift production. In the same way the requirements differ in a cycle operation and in a job shop production with buffers.

As soon as the existence of a condition monitoring is necessary for the machine tool operation the IPS^2 provider is interested in a cost efficient solution for sensors and evaluation electronics. The second objective of this chapter is the development and application of a cost effective wireless sensor network for distributed condition monitoring. This wireless sensor network enables the IPS^2 provider to upgrade machine tools with low effort.

10.3 Method for Ensuring the Availability of IPS^2

10.3.1 Concept

A method is specified by activities, roles, specification documents, techniques, and a meta model [2, 11]. In the framework of this chapter the method for ensuring the availability of IPS^2 shall be described by means of activities and tools. The sequence of activities considers the IPS^2 lifecycle which consists of the phases planning, development, implementation, operation, and resolution [9]. Figure 10.1 shows the assignment of activities and tools to the IPS^2 lifecycle phases.

In an availability-oriented IPS^2 business model the provider prices the solution according to a specific technical availability. Therefore, the effort for the maintenance activities has to be calculated at an early stage of the IPS^2 lifecycle. In the framework of this chapter a workflow management system has been used to model and simulate the maintenance costs according to the components' properties and their operation conditions over the entire lifecycle. In the case where the integration of a condition monitoring solution represents the optimal approach the machine tool components have to be extended in the implementation phase by the integration of additional hardware by means of sensors and software for the data acquisition and analysis. During the IPS^2 operation the periodic acquisition of sensor data allows the classification of the results and therefore the assessment of wear margin for a specific component. In the following chapter the activities of the method will be described in particular.

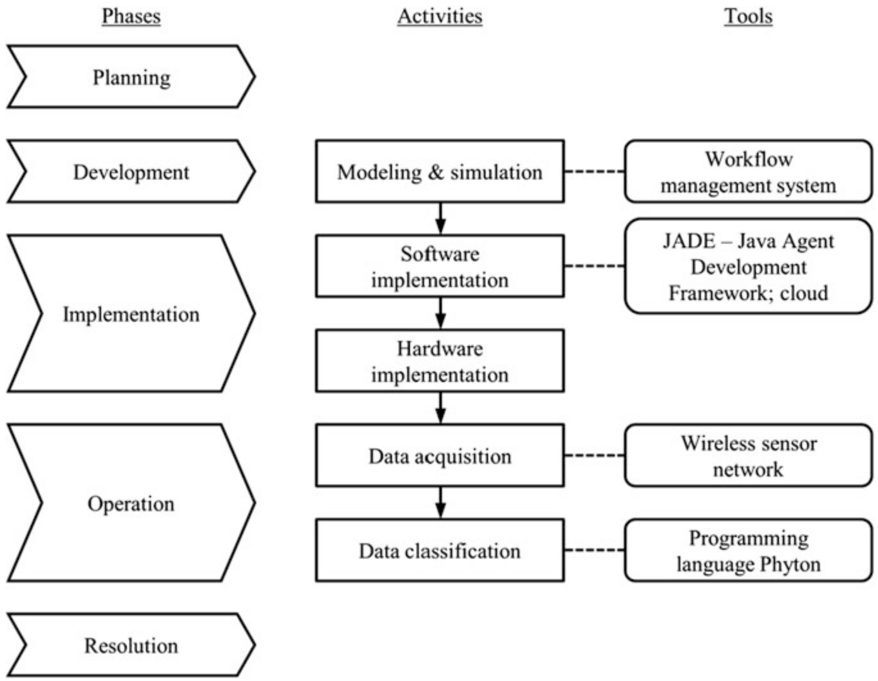


Fig. 10.1 Assignment of activities and tools to the IPS^2 lifecycle phases

10.3.2 Activities

10.3.2.1 Modeling and Simulation

Concrete activities shall be described by means of a ball screw spindle used in an NC-controlled centerless grinding machine feed axis. In the development phase the IPS^2 provider has to choose an appropriate concept to minimize the maintenance costs. This can be achieved by a reactive maintenance concept, the storage of spare parts on customer site, or an implemented condition monitoring system. According to the information of a machine tool manufacturer, the ball screw spindle exchange is necessary approximately every 5 years. As a consequence the ball screw spindle has to be exchanged twice regarding a machine tool's lifetime of 15 years. Figure 10.2 shows the generic business process for the exchange of components. This process model was created using the workflow management system IYOPRO developed by the company intellivate GmbH.

The process execution varies depending on the existence of a condition monitoring system or the storage of spare parts on customer site. If the condition monitoring is nonexistent, the service technician arrives at the customer site, analyzes the error, and checks the on-site availability of the necessary spare part. In cases where the appropriate spare part is not available a procurement activity is

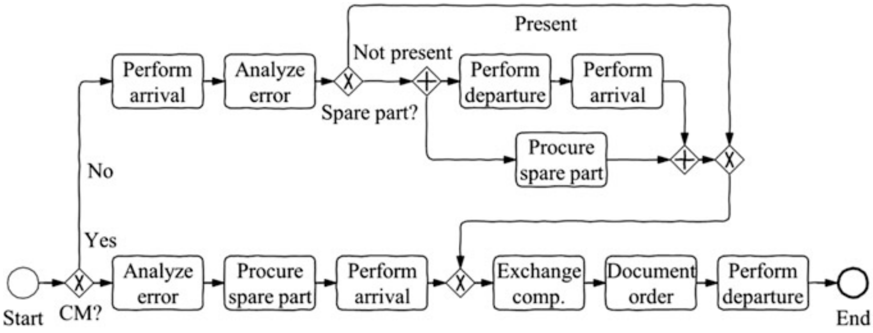


Fig. 10.2 Generic business process model for the exchange of components

Table 10.1 Durations of activities

Activity	Duration (h)			
	Reactive maintenance	Spare part	Wireless sensor network	Actual condition monitoring
Arrival	4	4	4	4
Spare part procurement	48	0	0	0
Analysis	1	1	0.25	0.25
Exchange	3	3	3	3
Documentation	0.5	0.5	0.17	0.17
Departure	4	4	4	4

necessary which may last several days and causes machine downtime. During that time the service technician departs and returns as soon as the spare part is available. After exchanging the component the service technician documents the order and departs from customer site. In the case where a condition monitoring is available the business process starts with the analysis of the error under consideration of appropriate sensor data and the procurement of spare parts. Subsequently, the service technician arrives at customer site and handles the order.

To identify the optimized maintenance concept the process model will be simulated applying the discrete event-based simulation (DES) method. IYOPRO allows the investigation of process models due to the implemented interface to the JAVA-based open-source bibliography DESMO-J [4]. Therefore, the process model has to be supplemented by durations of activities and costs of resources. For the durations of activities the four cases reactive maintenance, spare part, wireless sensor network, and actual condition monitoring were differentiated (Table 10.1).

Table 10.2 shows the costs of resources in the present process model. The costs for machine downtime differ with regard to the production type and the existence of buffers. The costs for a machine hour with buffers decrease in a three-shift production due to the optimized degree of utilization. Regarding a production type

Table 10.2 Costs of resources

Resources	Costs (€)			
	One-shift with buffer	One-shift without buffer	Three-shift with buffer	Three-shift without buffer
Machine downtime per hour	71.65	600	54.91	400
Service technician per hour	46.54	46.54	46.54	46.54
Travelling costs per trip (one-way)	100	100	100	100

without buffers the machine costs per hour are several times higher. The costs were estimated for the application scenario of an NC-controlled centerless grinding machine in a job shop production. Depending on the type of machine tools these costs can amount up to several thousand euros.

The process model was simulated for the four cases one-shift with buffer, one-shift without buffer, three-shift with buffer, and three-shift without buffer (Fig. 10.3). The reactive maintenance concept increases the maintenance costs in all four scenarios. This is due to the fact that the procurement of the necessary spare part may lead to a machine stop over several days.

Another interesting finding relates to the spare parts supply. In production scenarios with buffers between particular machine tools the storage of spare parts on customer site causes lower costs than the implementation of a condition monitoring according to the state of the art.

The implementation of the wireless sensor network for a proactive maintenance concept causes the lowest costs in all cases. This is primarily the result of the lower hardware and software costs compared to actual condition monitoring solutions. The development of the software and hardware will be described in the following chapters.

10.3.2.2 Software Implementation

The development of software tools ensures the interfaces between the components of the overall system (Fig. 10.4). The concept consists of the machine tool component, the wireless sensor network, which includes a MEMS vibration sensor, a minicomputer (Raspberry), the cloud (including services), and a smart device for remote access and visualization.

The data acquisition and processing is realized on the sensor node level. Furthermore, the classification of the features extracted from the diagnosis signal is transmitted to the cloud server. On the cloud server there are applications and services available for maintenance planning, remote condition monitoring, etc.

Via smart device the service technician or maintenance planner can receive the reports generated using a condition monitoring processing unit. This allows remote

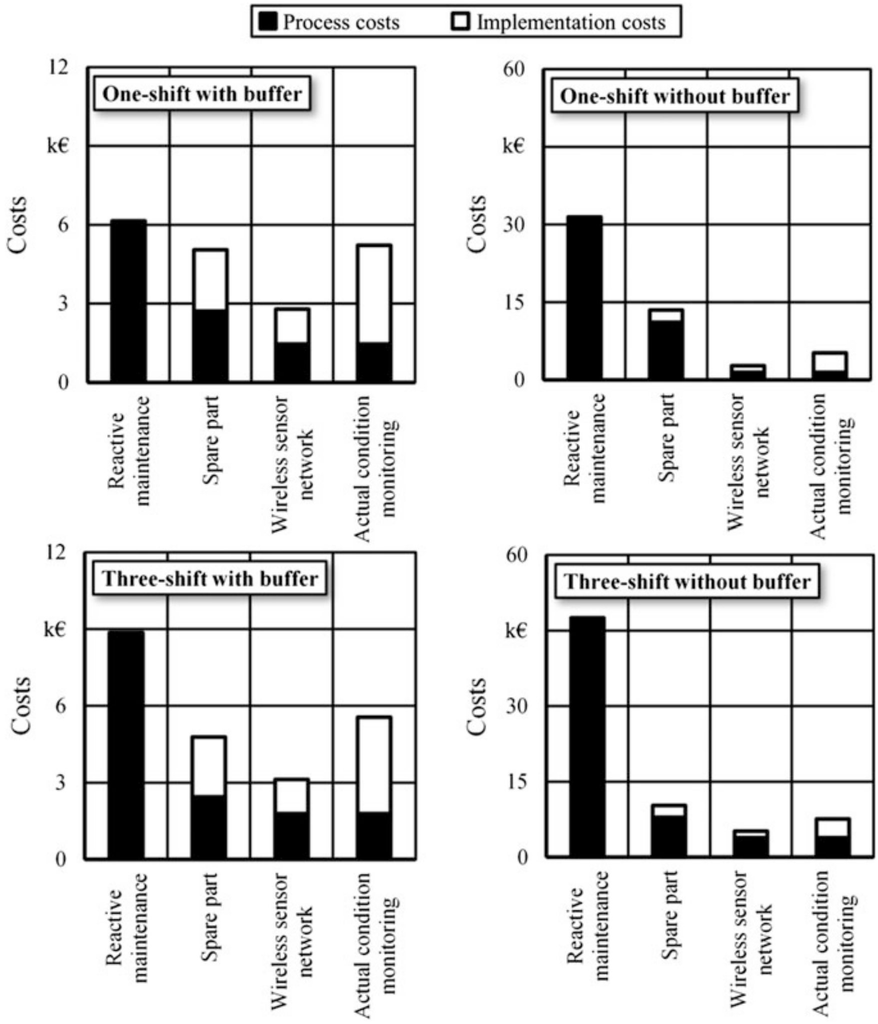


Fig. 10.3 Results of the simulation experiment

continuous monitoring of the production system and enables condition-based maintenance decision making of the system concerned.

The communication between these subsystems has been realized by the implementation of a software agent system. The Java Agent Development Framework (JADE) will be used as a middleware (Fig. 10.1). The internal communication between the software agents is performed through the asynchronous exchange of agent communication language (acl) messages. Two software agents interact by means of defined pattern. A communication always consists of an initiator and a responder. Initially, an agent demands another agent's readiness to execute a service

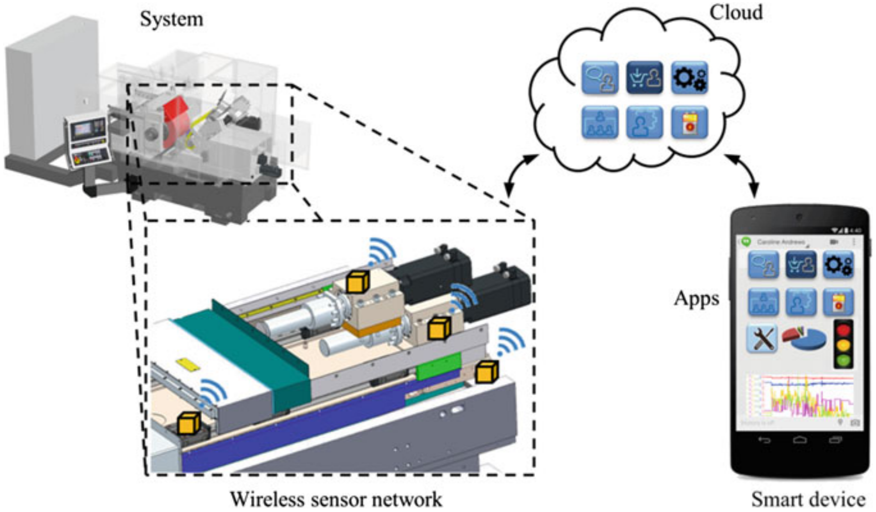


Fig. 10.4 Wireless sensor network on the production system

by a 'call-for-proposals' (cfp). As soon as the service is executable the second agent replies 'propose'. Afterwards, the initiator agent executes the service by sending the message 'accept-proposal'. After finishing the execution of the service the responder agent replies 'inform-done'. This general procedure will be performed until the overall process execution is finished. Figure 10.5 exemplarily shows a part of the communication between several agents for the execution of the feed axis test run.

Initially, the machine tool agent secures the readiness of the cloud for the collection of data. If the cloud agent agrees the request, the axis test run can be ordered. The directory facilitator (DF) agent, a directory which announces which agents are available on the platform, informs about the necessary services to execute the test run. The agents and services will be blocked for the duration of the execution. Subsequently, the machine tool agent communicates consecutively with the agents to particularly perform the activities of the axis test run. After finishing the test run the result data are submitted to the cloud. The comparison of the results with previous test runs allows the assessment of wear margin of the feed axis.

10.3.2.3 Hardware Implementation

The modifications of the components will be described by means of a feed axis test rig. It incorporates a top slide, an inside slide, and a machine bed. Furthermore, two asynchronous motors drive the ball screw spindles, which move the slides along the unloading axis X4, and the infeed axis and recessing axis X1 in steps of 0.2 μm. Figure 10.6 shows a schematic representation of the axis test rig.

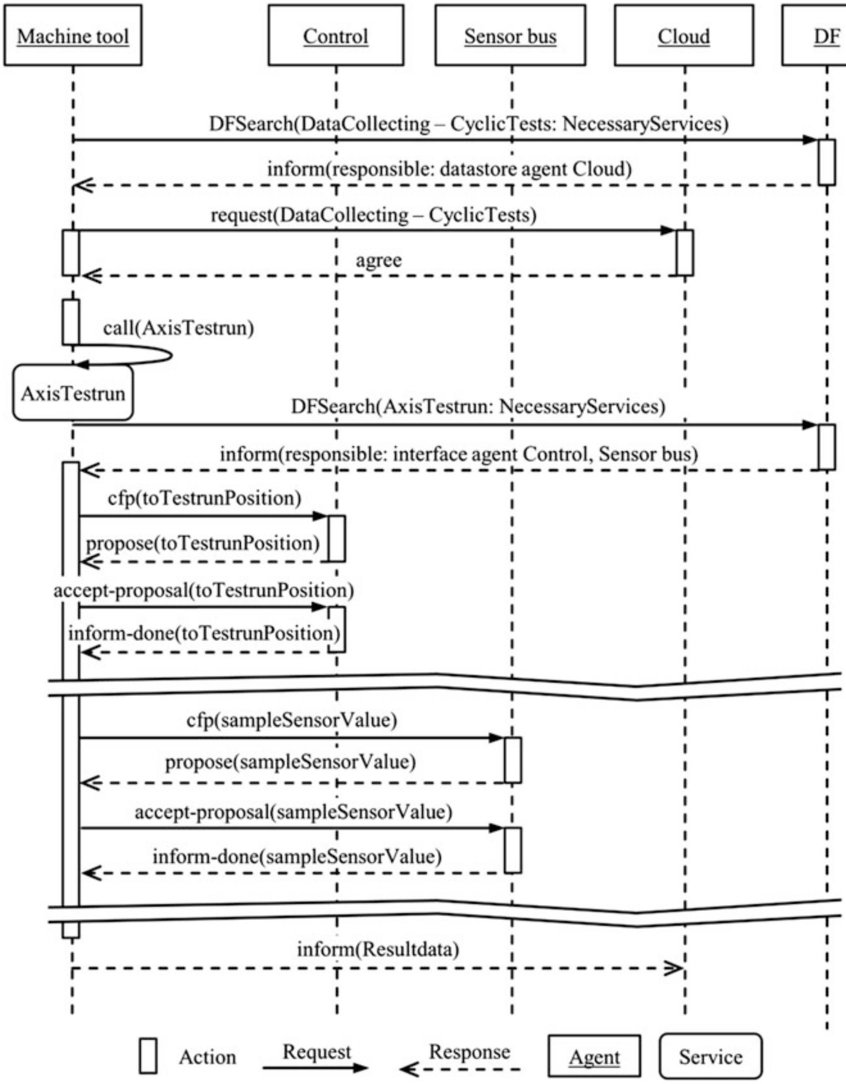


Fig. 10.5 Unified modeling language (UML) interaction diagram of the agent communication during the feed axis test run

The wireless sensor network is composed of four individual nodes which can act independently from each other. Each of these nodes is comprised of a MEMS temperature and vibration sensor. Furthermore, a Micro-Electro-Mechanical-Systems (MEMS) digital output motion sensor with 3-axes ‘nano’ accelerometer is used (Table 10.3).

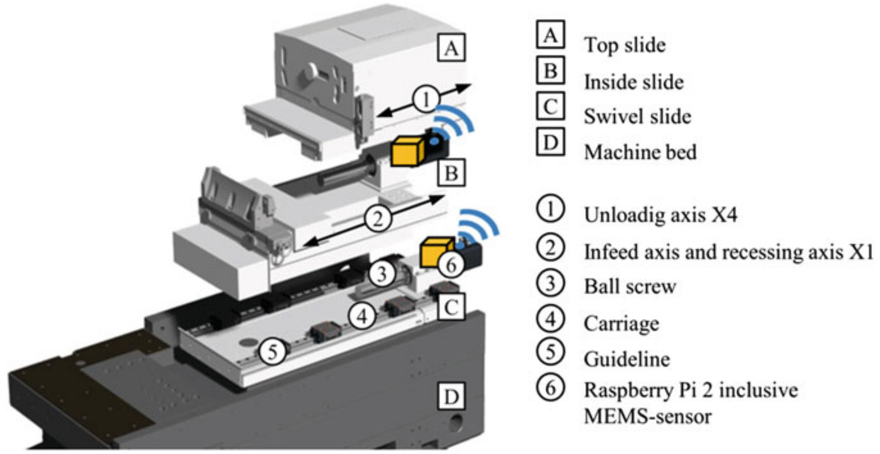


Fig. 10.6 Schematic representation of the axis test rig

Table 10.3 MEMS sensor LIS3DH specification

Feature	Value
Measuring accelerations with output data rates	1 Hz–5 kHz
Wide supply voltage	1.71–3.6 V
Ultra-low power mode consumption	<2 μA
Dynamically selectable full-scale	±2 g/±4 g/±8 g/±16 g
Data output	16 bit
Operating range	−40 to +85 °C
Shock survivability	10,000 g
Digital output interface	I2C/SPI

10.3.2.4 Data Acquisition

A design of experiments was developed for the acquisition of training data in order to evaluate the implemented algorithms. For this purpose, reproducible damages were created using the laser powder cladding method. A five-axes Trumpf TruLaser 7020 three dimensional laser cutting machine with various laser spot diameters was used to create a small and a heavy damage on the surface of the spindles (Fig. 10.7).

In addition, a fault was created on the needle roller/axial cylindrical roller bearing spindle. The acceleration of the spindles was measured on the axis test rig. They were conducted repeatedly under exactly the same conditions. The vibration was measured using a MEMS sensor with a sampling rate of 3,000 Hz. Figure 10.7 shows the collected vibration data at 1 and at 3 m/min for three different classes. It also shows the measured vibration data using MEMS sensor at the axis test rig at different operation speed.

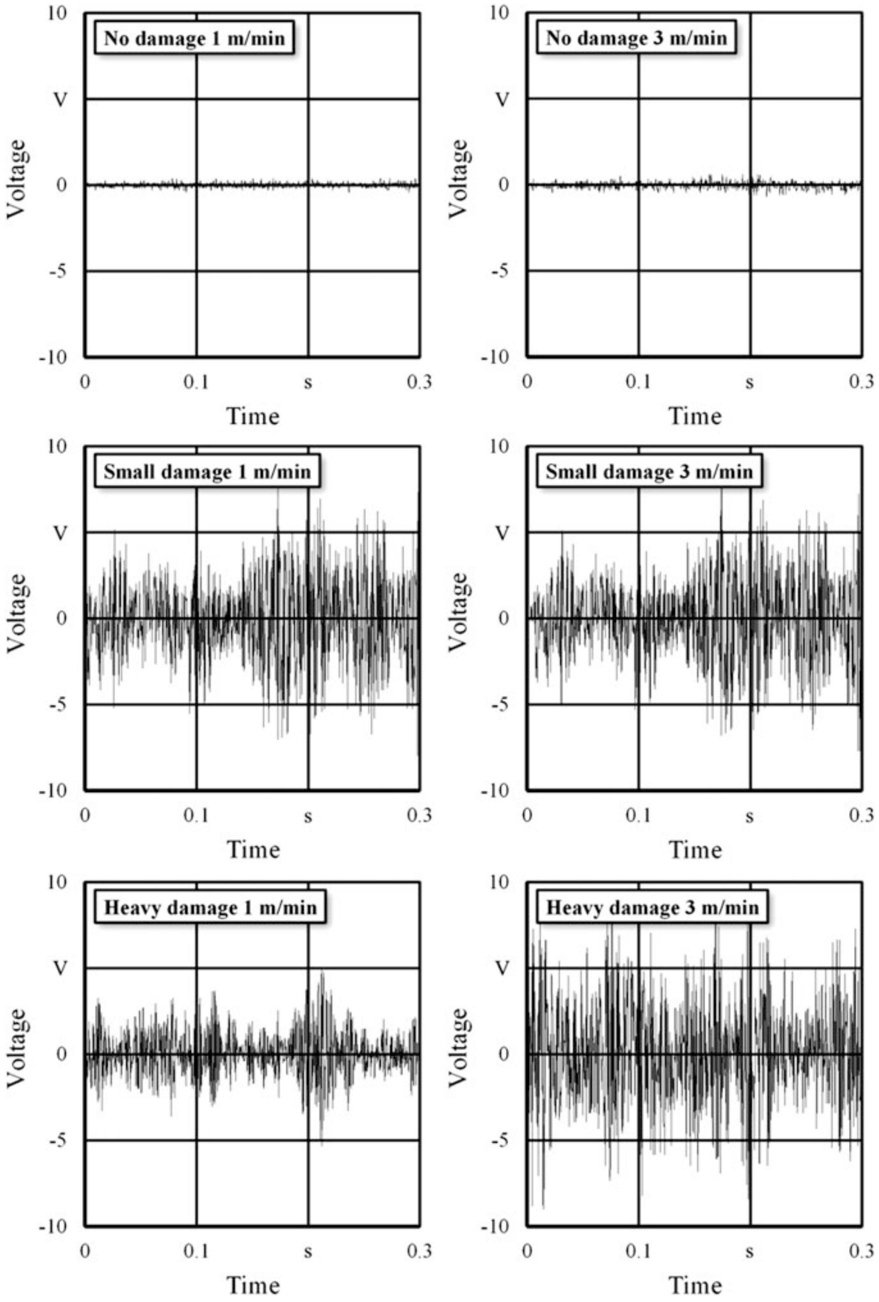


Fig. 10.7 Measured vibration data using MEMS sensor

10.3.2.5 Data Classification

The feature extraction using Python is realized on the Raspberry (Fig. 10.1). Python was selected to be used for this application because it is a fast programming language widely used in the field of data analysis using machine learning methods. To enable the feature calculation and classification on the Raspberry, extensions to the Python programming language such as Numpy, SciPy, Matplotlib, and Scikit-learn were used.

For the calculation of the selected features, such as statistical values (variance, mean, and kurtosis), the Scipy library is used. It includes the functions for the calculation of the statistical values.

For the classification of the features extracted, a support vector machine (SVM) algorithm is implemented [7]. SVM is a computational learning method based on the statistical learning theory. This function acts as an expert system which is based on the Vapnik-Chervonenkis (VC) Theory [6] and has recently emerged as a general mathematical framework for estimating dependencies from finite samples. VC theory combines fundamental concepts and self-consistent mathematical theory, well-defined formulation and principles related to learning. SVM is used successfully in many classification problems like text categorization, image classification, and bioinformatics.

For fault detection and diagnosis in industrial applications SVM is developed for recognizing patterns in the collected sensor data. These are classified to predefined fault condition of the considered component [3].

The most significant benefit of SVM is higher efficiency in high dimensional nonlinear classification problems while other statistical classifiers often fail in achieving such efficiency. The idea is to maximize the margin between hyper plane and the training examples. This can be done by finding the optimal hyper plane which has maximal margin.

After signal preprocessing, the statistical features are extracted. For the purpose of classification appropriate data has been selected. The implementation of the SVM algorithm on the Raspberry is done in Python. The feature extraction and SVM algorithm was previously implemented and tested in a MATLAB environment. Figure 10.8 represents selected features for the classification step. The well-defined areas in the diagram show that the features selected (mean, kurtosis, and variance) for the classification step of the data at 1 m/min are correctly chosen.

The classification accuracy using different test data at 1 and 3 m/min is illustrated in Fig. 10.9.

The graph shows that the features selected for the classification (Fig. 10.7) for no damage and heavy damage at 1 m/min is 100%. In contrast, the accuracy of SVM classification of the test data for small damages did not reach 100% for both speeds. To increase the accuracy for small damages other feature combinations has to be selected.

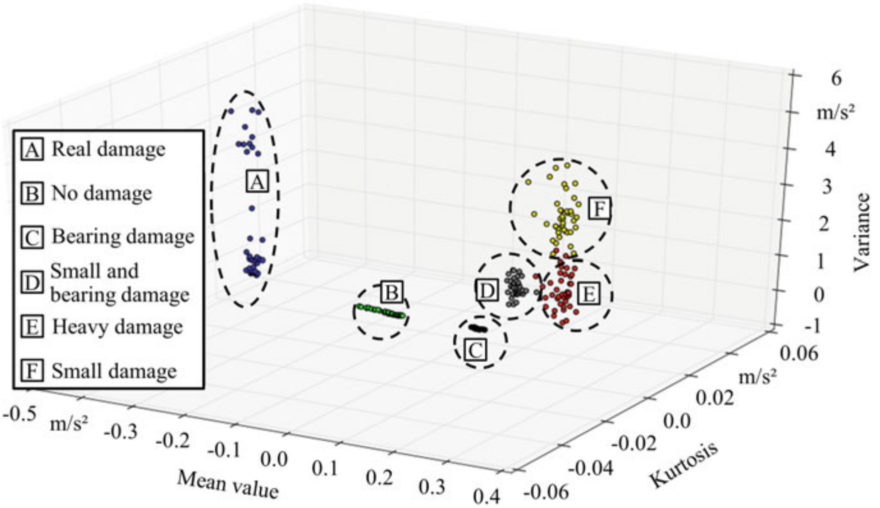
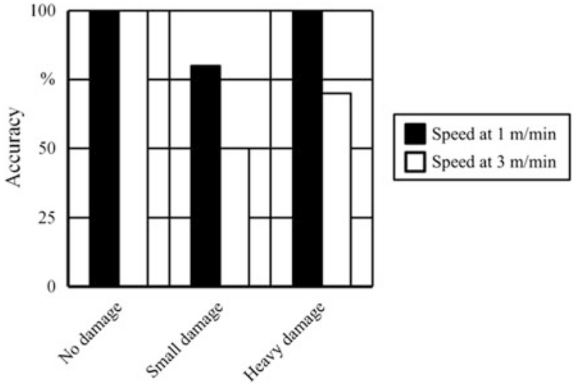


Fig. 10.8 Presentation of the suitable feature for the classification on the feature space

Fig. 10.9 SVM classification accuracy



10.4 Summary

This chapter presented a method for the customer individual modeling, simulation, and implementation of an appropriate MRO strategy for an availability-oriented IPS^2. Therefore, a generic business process for the exchange of worn components has been modeled, implemented, and simulated using the workflow management system IYOPRO developed by the company intellivate GmbH. The development of a simulation model allowed the comparison of different MRO strategies regarding the costs, i.e. occurred by a worn ball screw spindle of a feed axis. In this chapter the four cases reactive maintenance, spare part supply, and two condition monitoring approaches were compared and the most suitable concept for a grinding machine at the customer site was identified.

Furthermore, the development and implementation of a condition monitoring system, consisting of a wireless sensor network using Raspberry Pi 2 modules and MEMS vibrations sensors, has been described. The use of this network reduces the costs compared to industrial sensors and allows an easy adaptability to different specific applications due to decentralized data preprocessing, feature extraction, selection, and classification on the sensor node level.

The implementation of different classification and clustering algorithms on the sensor nodes will be addressed during further research activities. This will enable the condition monitoring steps (signal preprocessing, feature extraction, and classification) of the acquired data directly on the node level.

Acknowledgements We express our sincere thanks to the Deutsche Forschungsgemeinschaft (DFG) for funding this research within the collaborative research project SFB/TR 29 on Industrial Product-Service Systems—dynamic interdependency of products and services in the production area. Furthermore, we express our sincere thanks to the company intellivate GmbH for providing a license of IYOPRO Premium.

References

1. Baines TS, Lightfoot H, Steve E et al (2007) State-of-the-art in product service-systems. P I Mech Eng B-J Eng 221(10):1543–1552. doi:10.1243/09544054JEM858
2. Braun C, Wortmann F, Hafner M et al (2005) Method construction—a core approach to organizational engineering. In: Haddad H, Liebrock LM, Omicini A et al (eds) Proceedings of the 2005 ACM symposium on applied computing (SAC), Santa Fe, New Mexico, USA, March 13–17, 2005. ACM, New York, pp 1295–1299
3. Joachims T (2002) Learning to classify text using support vector machines. Springer, New York
4. Joschko P, Haan J, Janz T et al (2012) Business process simulation with IYOPRO and DESMO-J. In: Bruzzone AG, Buck W, Cayirci E et al (eds) Proceedings of the international workshop on applied modeling and simulation, Rende
5. Meier H, Roy R, Seliger G (2010) Industrial Product-Service Systems—IPS^2. CIRP Ann Manuf Technol 59(2):607–627. doi:10.1016/j.cirp.2010.05.004
6. Shawe-Taylor J, Cristianini N (2004) Kernel methods for pattern analysis. Cambridge University Press, Cambridge
7. Steinwart I, Christmann A (2008) Support vector machines. Springer, New York
8. Tukker A (2004) Eight types of product-service system: eight ways to sustainability? Experiences from SusProNet. Bus Strategy Environ 13(4):246–260. doi:10.1002/bse.414
9. Uhlmann E, Bochnig H (2012) PSS-CAD: an approach to the integrated product and service development. In: Shimomura Y, Kimita K (eds) The philosopher's stone for sustainability—Proceedings of the 4th CIRP international conference on industrial product-service systems, Tokyo, Japan, November 8th–9th, 2012. Springer, Heidelberg, pp 61–66
10. Vandermerwe S, Rada J (1988) Servitization of business: adding value by adding services. EMJ 6(4):314–324. doi:10.1016/0263-2373(88)90033-3
11. Winter R, Schelp J (2006) Reference modeling and method construction—a design science perspective. In: Haddad H (ed) Proceedings of the 2006 ACM symposium on applied computing (SAC), Dijon, France, April 23–27, 2006. ACM, New York, pp 1561–1562

Chapter 11
An Erlang-Coxian-Based Method for Modeling Accelerated Life Testing Data

Haitao Liao, Ye Zhang and Huairui Guo

Abstract Accelerated life testing (ALT) can be used to expedite failures of a product for predicting the product's reliability under the normal operating conditions. The resulting ALT data are often modeled by a probability distribution along with a life-stress relationship. However, if the selected probability distribution cannot adequately describe the underlying failure process, the resulting reliability prediction would be misleading. It would be quite valuable if the distribution providing an adequate fit to the ALT data can be determined automatically. This chapter provides a new analytical method to assist reliability engineers in this regard. Essentially, this method uses Erlang-Coxian (EC) distributions, which belong to a particular subset of phase-type distributions, to characterize ALT data. Such distributions are quite efficient for approximating many non-negative distributions, such as Weibull, lognormal and gamma. The advantage of this method is that the best fit to the ALT data can be obtained by gradually changing the model structure, i.e., the number of phases of the associated continuous-time Markov chain (CTMC). To facilitate the implementation of this method, two statistical inference approaches are provided. First, a mathematical programming approach is formulated to simultaneously match the moments of the EC-based ALT model to the empirical moments at the corresponding test stress levels. This approach resolves the feasibility issue of the method of moments. In addition, the maximum likelihood estimation approach is presented, which can easily handle different types of censoring in ALT. Both approaches are accompanied with a stopping criterion for determining the number of phases of the resulting CTMC. Moreover, nonparametric bootstrap method is used to construct the pointwise confidence interval for the resulting reliability estimates. Numerical examples for constant-stress ALT with Type-I and multiple censoring schemes are provided to illustrate the capability of the method in modeling ALT data.

H. Liao (✉) · Y. Zhang · H. Guo
Department of Industrial Engineering, The University of Arkansas,
Fayetteville, AR 72701, USA
e-mail: liao@uark.edu

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_11

Keywords Accelerated life testing · Phase-type distributions · Erlang-Coxian distributions · Method of moments · Maximum likelihood estimation

Notation

ALT	Accelerated life testing
AFT	Accelerated failure time
AIC	Akaike Information Criterion
Cdf	Cumulative distribution function
pdf	Probability density function
CTMC	Continuous-time Markov chain
PH	Phase-type
EC	Erlang-Coxian
EM	Expectation-maximization
LSE	Least square estimate
MLE	Maximum likelihood estimate
$F(tZ)$	Cdf of failure time under stress Z
$R(t; Z)$	Reliability function under stress Z
$f(t; Z)$	Pdf of failure time under stress Z
$h(t; Z)$	Hazard rate function under stress Z
$r(Z; \theta)$	Function of stress Z
$\boldsymbol{S}$	Subgenerator matrix
λ_j	The jth transition rate
p_j	The jth transition probability
π	Initial probability
M_l^i	The lth empirical moment under stress level i
τ_ι	Censoring time at stress level i
$\delta_{ij}^{[0,\tau_i]}$	Indicator function for the jth failure time under stress level i
$\varepsilon_{i,l}, s_{i,l}$	Excess and slack variables for the lth moment for stress level i
$w_{i,l}$	Weights assigned to the deviation from the lth moment for stress level i

11.1 Introduction

As technology advances, new products can be made quite reliable. For such a product, it is difficult, if not impossible, to observe failures in a short time period under the product's normal operating conditions. Accelerated life testing (ALT) has been widely used in industry as a viable tool for estimating the long-term reliability of such a product. The basic idea of ALT is to expose some units of the product to harsher-than-normal operating conditions to expedite failures. Based on the resulting failure time data due to acceleration, a statistical model is developed and used to extrapolate the product's long-term reliability under the normal operating conditions.

Nonparametric methods are practical choices for estimating the reliability of a product without assuming the underlying failure time distribution. The most popular ones are the Kaplan-Meier estimator and Breslow estimator [21]. However, it is difficult to extend these methods for developing ALT models which require the inclusion of various life-stress relationships for extrapolation in time and stresses. As a result, the most popular methods in modeling ALT data are to develop parametric models based on specified probability distributions with stress-dependent parameters. Such methods, if properly applied, are quite efficient, and the related statistical inference procedures, such as maximum likelihood estimation (MLE) and least squares estimation (LSE), have been extensively studied and made available to practitioners [9, 21, 23, 37].

Accelerated failure time (AFT) models are probably the most widely used parametric ALT models. Mathematically, an AFT model for the cumulative distribution function (Cdf) $F(t;Z)$ of failure time under a constant stress Z can be expressed as [3]:

$$F(t;Z) = 1 - R(t;Z) = F_0(r(Z;\underline{\theta})t), \tag{11.1}$$

where R(t; Z) is the corresponding reliability function, $F_0(\cdot)$ is the baseline Cdf, and $r(Z;\underline{\theta})$ is a deterministic function of Z. The model can also be expressed in terms of the corresponding hazard function as:

$$h(t;Z) = h_0(r(Z;\underline{\theta})t)r(Z;\underline{\theta}), \tag{11.2}$$

where $h_0(\cdot)$ is the corresponding baseline hazard function.

When using AFT models for reliability prediction, practitioners often face the challenge of choosing a probability distribution that provides an adequate fit to the collected ALT data. To determine the best model from several candidates, the likelihood values or residual plots (e.g., Cox-Snell residuals) of these ALT models can be considered. For relevant methods of goodness-of-fit tests, readers are referred to Bagdonavičius and Nikulin [3]. Regarding more general ALT models, Elsayed et al. [11] proposed the extended linear hazard regression model that is capable of modeling various ALT data and includes many ALT models as special cases. However, all these methods require the baseline Cdf's (or equivalently the baseline hazard functions) to be pre-specified. To assist engineers in modeling ALT data, a generic method using a collection of versatile distributions would be desirable for a wide range of engineering applications. It would be more attractive if the distribution providing the best fit to the ALT data can be determined adaptively.

Because the versatility of phase-type (PH) distributions naturally meets broad requirements for parametric modeling of failure time data, we propose a generic method using a specific and yet flexible subset of PH distributions, called Erlang-Coxian (EC) distributions [28], to model ALT data. Both mathematical programming and MLE approaches are developed for statistical inference.

The contribution of this generic method to the body of ALT literature is twofold. First, the use of EC distributions relaxes strong assumptions about the underlying failure time distributions in developing AFT models. Moreover, the method is able to achieve the best fit to the ALT data under a stopping criterion by adaptively adjusting the number of phases of the associated continuous-time Markov chain (CTMC).

The remainder of this chapter is organized as follows. Section 11.2 briefly reviews the literature on ALT models and PH distributions. Section 11.3 introduces the EC distributions and provides the mathematical formulation of the proposed EC-based ALT model. The mathematical programming and MLE approaches to the estimation of model parameters are presented in Sect. 11.4. Two simulation examples and a case study are provided in Sect. 11.5 to illustrate the capability of the EC-based ALT model for modeling ALT data. Finally, Sect. 11.6 concludes this chapter.

11.2 Literature Review

Among various probability distributions, the exponential distribution has been widely used in modeling ALT data. Typical examples include the models developed by Lawless and Singhal [16], Bai et al. [5], Bai and Chung [4], and Xiong [36]. The obvious limitation of the exponential distribution is that it can only be used for products with constant failure rates. To predict the reliability of a product with time-dependent failure rate, Nelson [22], Meeker [20], Bhattacharyya and Soejoeti [6], and Tang et al. [34] considered the Weibull distribution and different stress loadings used in ALT experiments. Moreover, Kielpinski and Nelson [15] developed ALT models based on the lognormal distribution and studied the optimum test plans. Doksum and Hóyland [7], and Onar and Padgett [27] used the inverse Gaussian distribution as the underlying distribution for ALT data. For testing of hypotheses for different distributions and thorough literature reviews of ALT models, readers are referred to Sethuraman and Singpurwalla [33], Escobar and Meeker [12], and Elsayed [10]. Despite these guidelines, choosing adequate distributions to fit ALT data is still a challenging task faced by engineers, and sometimes the underlying distributions are quite complex or even unknown.

It would be useful if a collection of probability distributions can be used to approximate those popular or even complex distributions arbitrarily closely so that choosing adequate distributions for ALT analysis can be circumvented. Indeed, PH distributions are quite versatile, which can be used for this purpose. Van Der Heijden [35] derived the lower and upper bounds on the third moment of a positive random variable when the squared coefficient of variation is between 0.5 and 1, and characterized the two-phase Coxian distributions that correspond to these bounds. Asmussen et al. [2] studied MLE for fitting PH distributions. An extended expectation-maximization (EM) algorithm was developed to minimize the information divergence in the density approximation case. Olsson [26] developed an EM algorithm to estimate the parameters of a PH distribution with fixed order from right-censored and interval-censored data. Riska et al. [30] developed a method to efficiently fit a long-tailed data set by partitioning the data set and fitting each partition into a hyperexponential distribution using an EM algorithm. Osogami and Harchol-Balter [28] provided closed-form solutions for mapping general distributions to quasi-minimal PH distributions (EC distributions) by matching their first three moments. In the field of reliability engineering, PH distributions have been used in modeling repairable systems. Sericola [32] proposed an algorithm to compute the interval availability of a system where a two-state semi-Markov process was used to model the system. An exponential distribution and a PH distribution were assumed for the operational and nonoperational state, respectively. Perez-Ocon and Montoro-Cazorla [29] studied the transient behavior of a system with operational and repair times following PH distributions. Moreover, Ruiz-Castro et al. [31] studied the availability and conditional failure probabilities for different types of failures of a multi-component system subject to internal and external repairable (or non-repairable) failures where the corresponding random times follow PH distributions.

The brief literature review indicates that PH distributions have not been used in modeling ALT data. To the best of our knowledge, this chapter demonstrates, for the first time, the potential of using PH distributions for ALT data analysis. Although the proposed model is developed based on a specific PH distribution, the concept and statistical estimation methods can be naturally extended to other types of PH models.

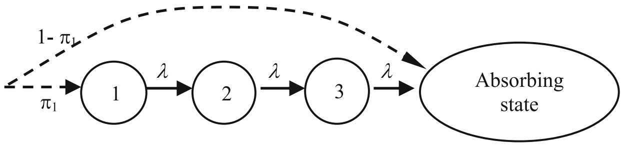
11.3 Mathematical Formulation of an EC-Based ALT Model

A PH distribution describes the time to absorption of a CTMC defined on a finite-state space [1]. Figure 11.1 shows two examples, where the time until absorption to the respective absorbing state in each CTMC can be characterized by a PH distribution. Essentially, such a CTMC with the specific structure can be described by an infinitesimal generator matrix:

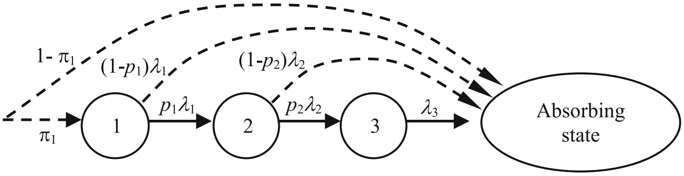
$$\mathbf{Q} = \begin{bmatrix} 0 & \mathbf{0} \\ \boldsymbol{S}^0 & \boldsymbol{S} \end{bmatrix}, \tag{11.3}$$

where $\boldsymbol{S}$ is the subgenerator matrix consisting of the transition rates among transient states and the ones for transitions into the absorbing state, and $\boldsymbol{S}^0 = -\boldsymbol{S}1$ in which $\mathbf{1} = (1, 1, \ldots, 1)'$ is a column vector of 1's of appropriate dimensions.

Let T be the time to absorption of a k-phase CTMC and $\boldsymbol{\pi} = (\pi_1, \pi_{\mathbf{2}}, \ldots, \pi_k)$ be the initial distribution of the CTMC. The Cdf of T can be expressed as:



(a) CTMC for a 3-phase Erlang distribution $E(3, \lambda)$



(b) CTMC for a 3-phase Coxian distribution $C(\lambda_1, \lambda_2, \lambda_3, p_1, p_2)$

Fig. 11.1 CTMCs whose absorption times define two different phase-type distributions

$$F(t) = 1 - R(t) = 1 - \boldsymbol{\pi}\ \exp\{\mathrm{t}\boldsymbol{S}\}\mathbf{1}, \quad t \geq 0 \tag{11.4}$$

where $\exp\{t\boldsymbol{S}\}$ is matrix exponential defined as:

$$\exp\{t\boldsymbol{S}\} = \sum_{k=0}^{\infty} \frac{1}{k!} (t\boldsymbol{S})^k. \tag{11.5}$$

The corresponding probability density function (pdf) is:

$$f(t) = \boldsymbol{\pi} \exp\{t\boldsymbol{S}\}\boldsymbol{S}^{\mathbf{0}}, t \geq 0 \tag{11.6}$$

and the hazard function is given by:

$$h(t) = f(t)/(1 - F(t)) = \boldsymbol{\pi} \exp\{t\boldsymbol{S}\}\boldsymbol{S}^0/(\boldsymbol{\pi} \exp\{t\boldsymbol{S}\}1), t \geq 0. \tag{11.7}$$

In addition, the lth moment of the distribution can be expressed as:

$$E[T^l] = (-1)^l l! \boldsymbol{\pi} \boldsymbol{S}^{-l}\mathbf{1}. \tag{11.8}$$

For examples, the CTMC in Fig. 11.1a results in a three-phase Erlang distribution $E(3, \lambda)$ with:

$$\boldsymbol{\pi} = (\pi_1, 0, 0) \text{ and } \boldsymbol{S} = \begin{bmatrix} -\lambda & \lambda & 0 \\ 0 & -\lambda & \lambda \\ 0 & 0 & -\lambda \end{bmatrix},$$

and the one in Fig. 11.1b gives a three-phase Coxian distribution $C(\lambda_1, \lambda_2, \lambda_3, p_1, p_2)$ with:

$$\boldsymbol{\pi} = (\pi_1, 0, 0) \text{ and } \boldsymbol{S} = \begin{bmatrix} -\lambda_1 & p_1\lambda_1 & 0 \\ 0 & -\lambda_2 & p_2\lambda_2 \\ 0 & 0 & -\lambda_3 \end{bmatrix},$$

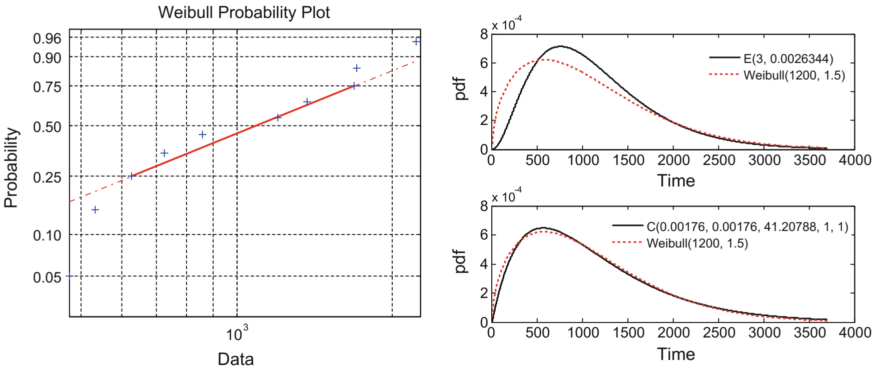
where $0 < p_i \leq 1$, $i = 1, 2$.

One of the most attractive properties of PH distributions is that the set of PH distributions is dense in the set of nonnegative distributions [24]. In other words, in theory, any nonnegative distribution can be approximated arbitrarily closely by a PH distribution. The only limitation is that PH distributions are light-tailed, thus may not be used as effective models for heavy-tailed distributions. Among different subsets of PH distributions, Erlang distributions and Coxian distributions are probably the most popular ones. They have been extensively studied in queueing theory and widely used in healthcare such as survival analysis and modeling the length of stay of patients in a hospital [18, 19].

An important aspect of mapping a general distribution to a PH distribution is to select the type and the number of phases of the corresponding CTMC. For example, for a data set (see Fig. 11.2a) generated from the Weibull(η = 1200, β = 1.5) distribution, a three-phase Erlang distribution $E(3, \widehat{\lambda} = 0.0026344)$ can be obtained through an MLE approach, and the corresponding log-likelihood value is 76.91. A three-phase Coxian:

$$C(\widehat{\lambda}_1 = 0.00176, \widehat{\lambda}_2 = 0.00176, \widehat{\lambda}_3 = 41.20788, \widehat{p}_1 = 1, \widehat{p}_2 = 1)$$

can also be obtained with log-likelihood value of 77.79. Obviously, the three-phase Coxian distribution, which turns out to be a hypoexponential distribution, provides a better fit (see Fig. 11.2b).



(a) Weibull plot of ten data points

(b) Results from fitted three-phase Erlang and Coxian distributions

Fig. 11.2 Example for Weibull (η = 1200, β = 1.5)

Before addressing the proposed EC-based ALT model, we first introduce the following definitions regarding the determination of a subset of PH distributions under certain estimation accuracy requirements [28].

Definition 1 Let $E[X^l]$ be the lth moment of random variable X with distribution G. The normalized lth moment m_l^G of X for l = 2,3 is defined as: $m_2^G = \frac{E[X^2]}{(E[X])^2}$ and $m_3^G = \frac{E[X^3]}{E[X]E[X^2]}$.

Definition 2 A distribution G is well represented by a distribution F if F and G agree on their first three moments. PH_3 refers to the distributions that are well represented by a PH distribution.

It is well known that a distribution G is in PH_3 if and only if its normalized moments satisfy $m_3^G > m_2^G > 1$ [14]. Since any nonnegative random variable with distribution G satisfies $m_3^G \geq m_2^G \geq 1$, almost all such distributions are in PH_3.

Osogami and Harchol-Balter [28] introduced EC distributions that are quite efficient for approximating PH_3 distributions. Figure 11.3 shows the CTMC for a k-phase EC distribution ($k \geq 3$), which consists of an Erlang $E(k-2, \lambda_1)$ with k-2 phases and a two-phase Coxian $C(\lambda_2, \lambda_3, p_c)$, for which:

$$\boldsymbol{\pi} = \underbrace{(\pi_1, 0, \ldots, 0)}_{} \text{ and } \boldsymbol{S} = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & & \cdots & & 0 \\ 0 & -\lambda_1 & \lambda_1 & 0 & & 0 & \\ \vdots & 0 & -\lambda_1 & \ddots & \ddots & & \vdots \\ & & 0 & \ddots & \lambda_1 & 0 & \\ \vdots & & & \ddots & -\lambda_1 & \lambda_1 & 0 \\ & 0 & & & 0 & -\lambda_2 & p_c\lambda_2 \\ 0 & \cdots & & \cdots & & 0 & -\lambda_3 \end{bmatrix}. \tag{11.9}$$

The idea of creating an EC distribution is that a two-phase Coxian distribution can well represent any distribution that has high second and third moments while an Erlang distribution has only two free parameters and has the least normalized second moment among all the PH distributions with a fixed number of phases. Therefore, such a k-phase EC distribution can represent probability distributions in

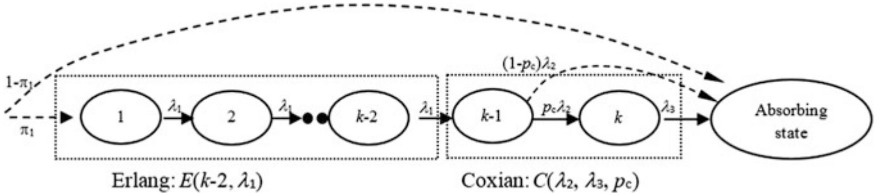


Fig. 11.3 CTMC whose absorption time defines a k-phase EC distribution

a wide range of variability using only a small number of phases. This is quite important in developing an ALT model, which needs additional parameters for quantifying the life-stress relationship.

Following Eqs. (11.1) and (11.4), the proposed EC-based ALT model (essentially an AFT model) for a product's failure time under stress level Z can be expressed as:

$$F(t;Z) = F_0(r(Z;\underline{\theta})t) = 1 - \boldsymbol{\pi}\ \exp\{r(Z;\underline{\theta})t\boldsymbol{S}\}\mathbf{1}, \quad t \geq 0, \tag{11.10}$$

where $\boldsymbol{S}$ is given in Eq. (11.9), and $\boldsymbol{\pi}$ is set to be $(1, 0, \ldots, 0)$ throughout this chapter.

The corresponding pdf is given by:

$$f(t;Z) = r(Z;\underline{\theta})\boldsymbol{\pi}\ \exp\{r(Z;\underline{\theta})t\boldsymbol{S}\}\ \boldsymbol{S}^{\mathbf{0}}, \quad t \geq 0, \tag{11.11}$$

and the hazard function is:

$$h(t;Z) = r(Z;\underline{\theta})\boldsymbol{\pi}\ \exp\{r(Z;\underline{\theta})t\boldsymbol{S}\}\boldsymbol{S}^0/(\boldsymbol{\pi}\exp\{r(Z;\underline{\theta})t\boldsymbol{S}\}\mathbf{1}), \quad t \geq 0. \tag{11.12}$$

In addition, the corresponding lth moment under stress level Z can be expressed as:

$$E[T_Z^l] = (-1)^l (r(Z;\underline{\theta}))^{-l} l! \boldsymbol{\pi} \boldsymbol{S}^{-l}\mathbf{1}. \tag{11.13}$$

To use such an EC-based ALT model in practice, it is necessary to determine the number of phases of the EC distribution adaptively based on the collected ALT data and estimate all the parameters. Effective solutions to these issues will make the proposed model more flexible in fitting ALT data than existing parametric models.

11.4 Statistical Inference Methods

This section addresses two statistical inference methods for estimating the parameters $\underline{\Theta} = [\lambda_1, \lambda_2, \lambda_3, p_c, \underline{\theta}]$ of an EC-based ALT model. In particular, a constant-stress ALT experiment with Type-I censoring is considered. Let τ_i be the censoring time under stress level Z_i, $i = 1, 2, \ldots, m$, n_i be the total number of units tested under Z_i, t_{ij} be the recorded failure/censoring time of unit j tested under Z_i.

11.4.1 Use of Mathematical Programming

For complete ALT data, the lth empirical moment of failure times under stress level Z_i can be calculated by: $M_l^i = (1/n_i)\sum_{j=1}^{n_i} t_{ij}^l$. For the case of Type-I censoring, the corresponding empirical moments can be calculated using the idea of data augmentation [25]:

$$M_l^i = \frac{1}{n_i}\left[\sum_{j=1}^{n_i}\left(\delta_{ij}^{[0,\tau_i]}t_{ij}^l + (1-\delta_{ij}^{[0,\tau_i]})\left(\tau_i + \int_0^\infty \frac{1-F(x+\tau_i;Z_i)}{1-F(\tau_i;Z_i)}dx\right)^l\right)\right], \tag{11.14}$$

where,

$\delta_{ij}^{[0,\tau_i]} = \{1, \text{ if } t_{ij} < \tau_i; 0, \text{ otherwise}\}$, $\int_0^\infty [1-F(x+\tau_i;Z_i)]/[1-F(\tau_i;Z_i)]dx$ is the mean residual life of a unit tested under Z_i and censored at time τ_i, and $F(\bullet;Z_i)$ is given by Eq. (11.10).

The traditional method of moments set several population moments equal to the corresponding empirical moments and solve those equations simultaneously for the model parameters. To overcome the feasibility issue of this method by allowing deviations, the mathematical formulation for determining the matching EC-based ALT model with the least number of phases (i.e., k) can be expressed as a mathematical programming problem (P_1):

$$\begin{array}{llll}
(\mathrm{P}_1) & \min\limits_{k} & k & \\
(\mathrm{SP}_1) & \underset{\substack{\underline{\theta},\lambda_1,\lambda_2,\lambda_3,p_c \\ \{s_{i,l},\varepsilon_{i,l}\},\, i\in\{1,\ldots,M\},\, l\in\{1,2,3\}}}{\min} & g_k = \sum\limits_{i=1}^{M}\sum\limits_{l=1}^{3}\left(\omega_{i,l}^{(1)}s_{i,l} + \omega_{i,l}^{(2)}\varepsilon_{i,l}\right) & \\
& \multicolumn{2}{c}{\text{Subject to}} & \\
& \frac{-(r(Z_i;\underline{\theta}))^{-1}1!\boldsymbol{\pi}\boldsymbol{S}^{-1}\mathbf{1}-M_1^i}{M_1^i} = \varepsilon_{i,1} - s_{i,1}, & \text{for} \quad i = 1,\ldots,m, & [\text{1st moment}] \\
& \frac{(r(Z_i;\underline{\theta}))^{-2}2!\boldsymbol{\pi}\boldsymbol{S}^{-2}\mathbf{1}-M_2^i}{M_2^i} = \varepsilon_{i,2} - s_{i,2}, & \text{for} \quad i = 1,\ldots,m, & [\text{2nd moment}] \\
& \frac{-(r(Z_i;\underline{\theta}))^{-3}3!\boldsymbol{\pi}\boldsymbol{S}^{-3}\mathbf{1}-M_3^i}{M_3^i} = \varepsilon_{i,3} - s_{i,3}, & \text{for} \quad i = 1,\ldots,m, & [\text{3rd moment}] \\
& \varepsilon_{i,1} \ge 0, \varepsilon_{i,2} \ge 0, \varepsilon_{i,3} \ge 0, & \text{for} \quad i = 1,\ldots,m, & \\
& s_{i,1} \ge 0, s_{i,2} \ge 0, s_{i,3} \ge 0, & \text{for} \quad i = 1,\ldots,m, & \\
& \multicolumn{2}{c}{\lambda_1 \ge 0, \lambda_2 \ge 0, \lambda_3 \ge 0, p_c \ge 0,} & \\
& \multicolumn{2}{c}{\text{Subject to}} & \\
& \multicolumn{2}{c}{g_k \le \widetilde{\mathrm{E}},} & \\
& \multicolumn{2}{c}{k \in \{3,4,\ldots\},} &
\end{array} \tag{11.15}$$

where $\varepsilon_{i,l}$ is the “excess” variable in the lth moment for stress level Z_i, $s_{i,l}$ is the corresponding slack variable, $\omega_{i,l}^{(1)}$ and $\omega_{i,l}^{(2)}$ are the weights assigned to the deviation from the lth moment (the conceived importance of the deviation), and $\tilde{\mathrm{E}}$ is the pre-specified overall tolerance.

This mathematical programming problem consists of two stages. For a given value of k, a sub-problem SP_1 is solved to obtain the minimal value of g_k, which corresponds to the best matching EC-based ALT model with the k phases. The procedure is continued until the first value of k and the corresponding EC-based ALT model are found, which satisfies the overall tolerance constraint. Note that the size of matrix $\boldsymbol{S}$ increases as the value of k increases, which only increases the number of phases in Erlang $E(k\text{-}2, \lambda_1)$ without increasing the number of parameters. Despite the simplicity in terms of formulation, the challenge of using this approach is that extensive computational effort is required in finding the solution that satisfies those nonlinear constraints. Moreover, this approach may not result in the desirable optimality properties of MLE.

11.4.2 Use of MLE Approach

The MLE approach has been widely used in modeling ALT because of its capability of handling different types of censoring. In particular, for ALT data with Type-I censoring, the log-likelihood function lnL can be expressed as:

$$\begin{aligned} ln\,L(\underline{\Theta} &= [\lambda_1, \lambda_2, \lambda_3, p_c, \underline{\theta}]) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[\delta_{ij}^{[0,\tau_i]}\ln\big(f\left(t_{ij}; Z_i\right)\big) + (1-\delta_{ij}^{[0,\tau_i]})\ln\big(1 - F(t_{ij}; Z_i)\big)\right] \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta_{ij}^{[0,\tau_i]}\ln\big(r(Z_i;\underline{\theta})\boldsymbol{\pi}\exp\{r(Z_i;\underline{\theta})t_{ij}\boldsymbol{S}\}\boldsymbol{S^0}\big) \\ &\quad + (1-\delta_{ij}^{[0,\tau_i]})\ln\big(\boldsymbol{\pi}\exp\{r(Z_i;\underline{\theta})t_{ij}\boldsymbol{S}\}\boldsymbol{1}\big) \end{aligned} \tag{11.16}$$

For a case with random censoring, the log-likelihood function can be obtained by replacing $\delta_{ij}^{[0,\tau_i]}$ by $\delta_{ij} = \{1,\ \text{if } t_{ij}\text{is a failure time}; 0,\ \text{otherwise}\}$ for each t_{ij}. The MLEs of the model parameters can be obtained by maximizing the log-likelihood function. In practice, different optimization algorithms, such as Quasi-Newton and Nelder-Mead algorithms, can be used.

A practical issue is to determine the number of phases in the EC-based ALT model. For such model selection problems under the framework of MLE, the Akaike Information Criterion (AIC) has been widely used. The AIC for an EC-based ALT model can be expressed as:

$$AIC = 2q - 2lnL(\lambda_1, \lambda_2, \lambda_3, p_c, \underline{\theta}), \tag{11.17}$$

where q is the number of parameters which is the same for EC-based ALT models with different numbers of phases when the number of parameters in $\underline{\theta}$ is fixed. As a result, using AIC in determining the best EC-based ALT model is equivalent to comparing the log-likelihood values of those candidate models. Note that the likelihood-ratio test is widely used for testing nested models, which is not appropriate in determining the number of phases in EC-based ALT models because different EC-based ALT models with different numbers of phases essentially have the same number of parameters.

11.4.3 Method for Constructing Confidence Intervals

To quantify the uncertainty in parameter estimates $\underline{\hat{\Theta}}$ and the consequent reliability estimates, a nonparametric bootstrap method is proposed to construct the interested confidence intervals. This is a sample-reuse method, which can be used when no efficient alternatives are both tractable and sufficiently accurate. For EC-based ALT models, this may be the case.

In this chapter, the bootstrap procedure is used to construct pointwise confidence intervals for reliability function (or Cdf). This procedure is developed based on the method described by Efron and Tibshirani [8]. The detailed steps are as follows:

1. Each sample (say bootstrap sample j) consisting of n_i data points for each stress level Z_i is obtained by sampling, with replacement, from the original ALT data.
2. Parameters $\underline{\hat{\Theta}}$ are estimated based on each bootstrap sample. For the mathematical programming approach, the new bootstrap estimates $\underline{\hat{\Theta}}_j$ are calculated by solving the sub-problem SP_1 in Eq. (11.15). For the MLE approach, $\underline{\hat{\Theta}}_j$ is obtained by maximizing Eq. (11.16).
3. Compute the corresponding reliability function using Eq. (11.10), given the bootstrap estimate $\underline{\hat{\Theta}}$
4. Repeat Steps 1–3 for B (say 5000) times to obtain a set of reliability estimates at time t as:

$$\{\widehat{R_1}(t), \widehat{R_2}(t), \cdots, \widehat{R_B}(t)\}.$$

5. Sort the set in increasing order for each desired time to give:

$$\{\widehat{R_{[1]}}(t), \widehat{R_{[2]}}(t), \cdots, \widehat{R_{[B]}}(t)\}.$$

6. Determine the lower and upper bounds of pointwise 100(1-α)% confidence interval for reliability function as:

$$[\widehat{R_{[v]}}(t), \widehat{R_{[u]}}(t)],$$

where $v = [\alpha B/2]^+$ and $u = [(1-\alpha/2)B]^+$.

11.5 Numerical Examples

In this section, we provide three examples to illustrate the capability of EC-based ALT models in analyzing ALT data. In the first two examples, complete and Type-I censored ALT data are simulated respectively from two popular parametric ALT models: (1) Inverse-Power-Law-Weibull model, and (2) Arrhenius-Lognormal model. The corresponding EC-based ALT models are developed using the mathematical programming method, and their estimation accuracy is studied. Afterwards, a case study is presented to demonstrate the use of the proposed EC-based ALT model and MLE method in predicting the reliability of a type of miniature lamps under the use condition based on randomly censored ALT data.

11.5.1 Complete Data from an Inverse-Power-Law-Weibull ALT Model

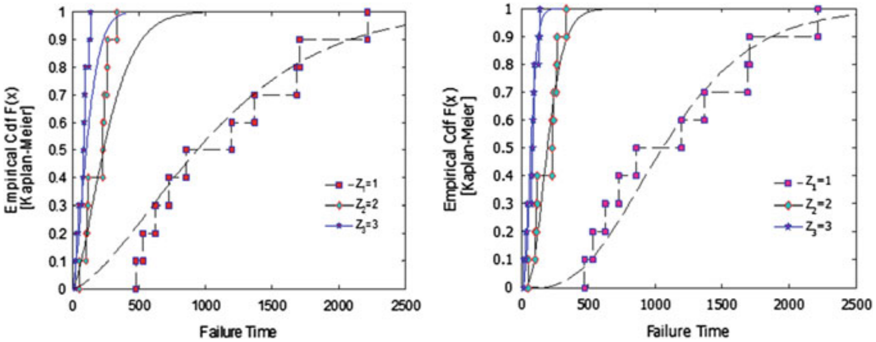
In the first example, a set of ALT data is generated from an inverse-power-law-Weibull ALT model with Cdf: $F(t; Z_i) = 1 - \exp(-(t/\eta(Z_i))^{\beta})$, where the shape parameter $\beta = 1.5$ and the scale parameter depends on stress level Z_i in the form of $\eta(Z_i) = 1200Z_i^{-2}$. This ALT model is a special case of AFT model. Three stress levels ($m = 3$): $Z_1 = 1$, $Z_2 = 2$, and $Z_3 = 3$, are considered in the hypothetic ALT experiment. At each stress level, ten units are tested to failure resulting in complete ALT data. Table 11.1 shows the data and the empirical moments. Figure 11.4a shows the empirical Cdf's for the simulated data and the Cdf's of the presumed ALT model for visual comparison. One can see that the empirical Cdf's significantly differ from the Cdf's of the presumed ALT model because of the randomness and small sample size of the ALT experiment. Without assuming a Weibull-based ALT model, it is more reasonable to develop an ALT model in a generic way that matches the ALT data as much as possible.

To estimate the model parameters of the EC-based ALT model through the mathematical programming method, we solve the nonlinear programming problem P_1, where i = 1, 2, 3, and the corresponding empirical moments:

$$M_l^i = (1/n_i)\sum_{j=1}^{n_i} t_{ij}^l,\ l = 1,\ 2,\ 3,$$

Table 11.1 ALT data generated from $F(t;Z_i) = 1 - exp\ (-(t/(1200Z_i^{-2}))^{1.5})$

Stress level	Failure times in hours					Empirical moments M_l^t		
						1st	2nd	3rd
$Z_1 = 1$	473.9	531.5	624.9	724.4	856.8	11318.8	1613526.0	2645512056.1
	1198.7	1367.3	1686.6	1706.3	2217.3			
$Z_2 = 2$	52.5	103.1	112.5	120.3	230.1	195.1	45529.2	11686756.3
	231.1	241.0	259.8	265.0	335.6			
$Z_3 = 3$	21.1	34.6	46.9	77.1	81.1	80.9	7859.4	844147.5
	87.6	94.6	97.9	130.8	137.5			



(a) Empirical Cdf's vs. Cdf's of the presumed model

(b) Empirical Cdf's vs. Cdf estimates using EC-based model

Fig. 11.4 4 Plots for ALT data generated from: $F(t;\, Z_i) = 1 - \exp\,(-t(/(1200Z_i^{-2}))^{1.5})$

Table 11.2 Deviations of the EC-based ALT model from the first three empirical moments

Stress level	Percentage deviations from the empirical moments		
	1st (%)	2nd (%)	3rd (%)
$Z_1 = 1$	0.00	–1.17	0.00
$Z_2 = 2$	7.38	18.54	40.94
$Z_3 = 3$	–3.84	–5.26	0.00

calculated for the *i*th stress level are given in the last three columns of Table 11.1, the weights assigned to the deviation from the *l*th moment are:

$$\begin{aligned}
\omega_{1,1}^{(1)} = \omega_{2,1}^{(1)} = \omega_{3,1}^{(1)} = \omega_{1,1}^{(2)} = \omega_{2,1}^{(2)} = \omega_{3,1}^{(2)} = 1,\\
\omega_{1,2}^{(1)} = \omega_{2,2}^{(1)} = \omega_{3,2}^{(1)} = \omega_{1,2}^{(2)} = \omega_{2,2}^{(2)} = \omega_{3,2}^{(2)} = 0.5,\\
\omega_{1,3}^{(1)} = \omega_{2,3}^{(1)} = \omega_{3,3}^{(1)} = \omega_{1,3}^{(2)} = \omega_{2,3}^{(2)} = \omega_{3,3}^{(2)} = 0.25,
\end{aligned}$$

respectively, and the tolerance level $\tilde{\mathrm{E}}$ is set to be 0.35. Moreover, the life-stress relationship is assumed to be $r(Z_i; \underline{\theta} = [\alpha_0, \alpha_1]) = \alpha_0 Z_i^{\alpha_1}$. Solving this problem using the interior-point algorithm yields parameter estimates:

$$\left[\widehat{\lambda_1}, \widehat{\lambda_2}, \widehat{\lambda_3}, \widehat{p_c}, \widehat{\alpha_0}, \widehat{\alpha_1}\right] = [41.520, 41.657, 41.504, 0.584, 0.0000969, 2.443],$$

for which $k = 5$ and $g_5^* = 0.3394 < \tilde{\mathrm{E}}$. The deviations of the resulting EC-based ALT model in the first three moments under different stress levels are given in Table 11.2. Each percentage error is calculated by: (moment of the EC-based ALT model - empirical moment)/empirical moment. One can see that the mathematical programming method provides quite accurate moment matching results. Figure 11.4b compares the empirical Cdf's for the simulated data and the Cdf's

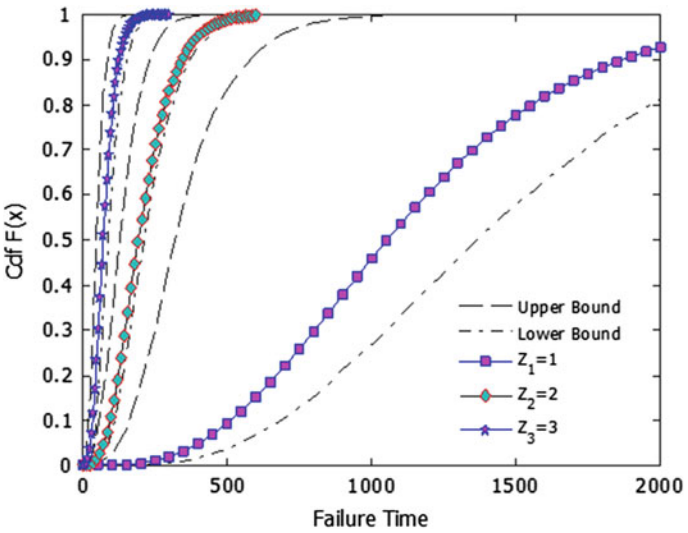


Fig. 11.5 Cdf estimates using the EC-based model and 95% bootstrap confidence intervals

estimated using the EC-based ALT model. It is clear that the Cdf estimates match the empirical Cdf's very well. To quantify the estimation uncertainty, the 95% confidence intervals of the Cdf's estimated using the EC-based ALT model (see Fig. 11.5) are also calculated using the bootstrap method addressed in Sect. 11.4.3.

11.5.2 Type-I Censored Data from an Arrhenius-Lognormal ALT Model

A set of censored ALT data is generated from an Arrhenius-lognormal ALT model $F(t; Z_i) = \Phi((\ln t - u(Z_i))/\sigma)$, where $\sigma = 1.5$ and the location parameter u depends on stress level Z_i in the form of $u(Z_i) = \ln(50) + 5/Z_i$. Again, three stress levels: $Z_1 = 1$, $Z_2 = 2$, and $Z_3 = 3$ are considered with ten units for each level. The censoring times at these stress levels are 80000, 6000, and 3000, respectively. Table 11.3 shows the simulated ALT data.

To develop an EC-based ALT model using the mathematical programming method, the weights assigned to the deviation from the lth empirical moment are again set to be:

$$\omega_{1,1}^{(1)} = \omega_{2,1}^{(1)} = \omega_{3,1}^{(1)} = \omega_{1,1}^{(2)} = \omega_{2,1}^{(2)} = \omega_{3,1}^{(2)} = 1,$$
$$\omega_{1,2}^{(1)} = \omega_{2,2}^{(1)} = \omega_{3,2}^{(1)} = \omega_{1,2}^{(2)} = \omega_{2,2}^{(2)} = \omega_{3,2}^{(2)} = 0.5,$$
$$\omega_{1,3}^{(1)} = \omega_{2,3}^{(1)} = \omega_{3,3}^{(1)} = \omega_{1,3}^{(2)} = \omega_{2,3}^{(2)} = \omega_{3,3}^{(2)} = 0.25,$$

Table 11.3 Censored ALT data generated from $F(t; Z_i) = \Phi((\ln t - \ln(50) - 5/Z_i)/1.5)$

Stress level	Failure times in hours ("+": censored unit)				
$Z_1 = 1$	2267.9	4758.9	10341.6	11700.3	
	17300.1	24762.7	79694.3	80000+	
	80000+	80000+			
$Z_2 = 2$	84.1	376.2	403.9	584.5	915.8
	1642.1	2435.5	5259.2	5689.4	6000+
$Z_3 = 3$	102.5	125.0	168.7	287.1	334.9
	338.1	775.4	1967.7	2971.1	3000+

Table 11.4 Deviations of the EC-based ALT model from the first three empirical moments

Stress level	Empirical moments $\boldsymbol{M}_l^t$			Percentage deviations from $\boldsymbol{M}_l^t$		
	1st	2nd	3rd	1st (%)	2nd (%)	3rd (%)
$Z_3 = 1$	1.147×10^5	4.155×10^{10}	1.784×10^{16}	18.33	–1.63	–10.07
$Z_2 = 2$	2.817×10^3	1.862×10^7	1.601×10^{11}	17.95	–9.27	–27.4
$Z_3 = 3$	1.349×10^3	5.488×10^6	2.991×10^{10}	3.50	–38.85	–63.01

respectively, and the tolerance level $\tilde{E}$ is set to be 0.95, which is determined through trial and error. Moreover, the life-stress relationship is assumed to be $r(Z_i; \underline{\theta} = [\alpha_0, \alpha_1]) = \alpha_0 \exp(-\alpha_1 Z_i^{-1})$. Unlike in the first example, the empirical moments in this example cannot be calculated directly from the data. Instead, Eq. (11.14) needs to be used, which depends on the model parameters to be estimated. By solving the corresponding nonlinear programming problem similar to Eq. (11.15), the parameters of the EC-based ALT model are estimated as: $[\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\lambda}_3, \widehat{p}_c, \widehat{\alpha}_0, \widehat{\alpha}_1] = [15.798, 209.5, 1799.7, 0.382, 0.0013, 5.895]$ for which $k = 5$ and $g_7^* = 0.9357 < \tilde{E}$. Table 11.4 shows the empirical moments calculated using Eq. (11.14) for the three stress levels and the deviations of the resulting EC-based ALT model in the corresponding moments. One can see that the deviations are well balanced across different stress levels as well as the corresponding moments.

11.5.3 Case Study

ALT has been widely used in evaluating the reliability of microelectronics [13, 38]. In this section, the ALT data reported by Liao and Elsayed [17] is used to illustrate the use of the proposed ALT model in practice. The purpose of this ALT experiment is to estimate the reliability of a type of miniature lamps under the use condition: 2 V. The highest operating voltage of the lamp is 5 V. It is well known

Table 11.5 ALT data of miniature lamps [17]

Stress level (V)	Failure times in hours ("+": censored unit)										
5	20.5	22.3	23.2	24.7	26	34.1	39.6	41.8	43.6	44.9	47.7
	61.6	62.1	65.5	70.8	87.8	118.3	120.1	145.4	157.4	180.9	187.7
	204	206.7	213.9	215.2	218.7	254.1	262.6	293	304	313.7	314.1
	317.9	337.7	430.2								
3.5	37.8	43.6	51.1	58.6	65.5	65.9	75.6	82.5	88.1	89	106.6
	113.1	121.1	121.5	128.3	151.8	171.7	181	202.7	211.7	230.7	249.9
	275.6	285	296.2	358.5	379.8	434.5	493.1	506.4	561.1	570	577.7
	876.3	890+	890+	890+	**922**	941+	941+				
2	223.1	254	316.7	560.2	679	737	894.4	930.5+	930.5+	930.5+	
	930.5+	930.5+	930.5+	930.5+	930.5+	930.5+	930.5+	930.5+	930.5+	930.5+	
	930.5+	930.5+	930.5+	930.5+							

that the coil temperature of an incandescent lamp during operation is mainly due to the electric current.

Three constant voltage levels were utilized in the experiment: 5, 3.5, and 2 V. Table 11.5 gives the observed failure times and censoring times under the three stress levels. It is worth pointing out that the test under 3.5 V was randomly censored and the one under 2 V was Type-I censored. To avoid making an assumption on the underlying distribution, such as Weibull and Lognormal, we use the proposed EC-based ALT model to predict the reliability of this type of miniature lamps.

To facilitate data analysis, we standardize the stress levels by defining $Z_i = [V_i - V_0]/[V_H - V_0]$, where $V_0 = 2$ V and $V_H = 5$ V. As a result, we have: $Z_1 = 1$, $Z_2 = 0.5$, and $Z_3 = 0$. The life-stress relationship is assumed to be characterized by:

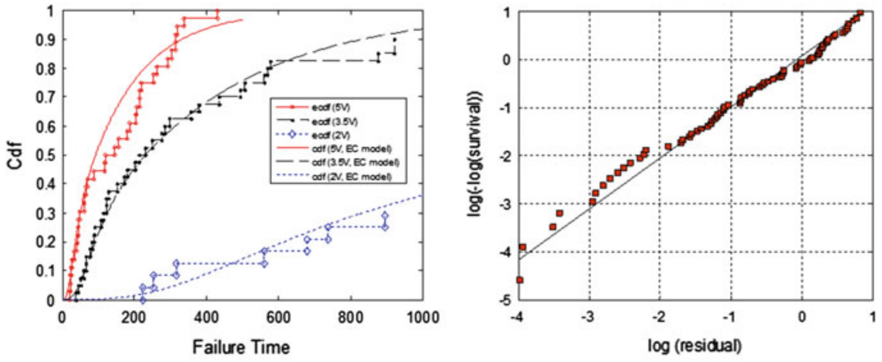
$$r(Z_i; \underline{\theta}) = \exp(\alpha_0 Z_i^{\alpha_1}). \tag{11.18}$$

Because the data set contains different types of censoring times, we use the MLE method introduced in Sect. 11.4.2 for statistical inference. Table 11.6 shows the MLEs of parameters for different EC-based ALT models with different numbers of phases. By comparing the log-likelihood values, the EC-based ALT model with $k = 7$ phases is selected after balancing the prediction accuracy and the complexity of the models

Figure 11.6 illustrates the statistical fittings of the resulting EC-based ALT model for the three test stress levels. The empirical Cdf (ecdf using Kaplan-Meier method) of the lamp under each voltage level is presented in Fig. 11.6a. Compared to the corresponding empirical Cdf's, this model exhibits satisfactory prediction capability. This can also be verified by examining the Cox-Snell residual plot presented in Fig. 11.6b, where the residuals can be easily tested against the

Table 11.6 MLEs of parameters for different EC-based ALT models with different numbers of phases (k)

Values of k	MLEs of parameters	Log-likelihood *lnL*
3: $E(1, \lambda_1)$ and $C(\lambda_2, \lambda_3, p_c)$	$\alpha_0 = 2.9091$; $\alpha_1 = 0.5762$; $\lambda_1 = 0.0026$; $\lambda_2 = 0.0026$; $\lambda_3 = 0.0003$; $p_c = 0.6730$;	–518.5038
4: $E\{2, \lambda_1)$ and $C(\lambda_2, \lambda_3, p_c)$	$\alpha_0 = 2.8807$; $\alpha_1 = 0.5730$; $\lambda_1 = 0.0045$; $\lambda_2 = 0.0045$; $\lambda_3 = 0.0003$; $p_c = 0.6980$;	–516.4058
5: $E(3, \lambda_1)$ and $C(\lambda_2, \lambda_3, p_c)$	$\alpha_0 = 2.8182$; $\alpha_1 = 0.5693$; $\lambda_1 = 0.0068$; $\lambda_2 = 0.0068$; $\lambda_3 = 0.0004$; $p_c = 0.7269$;	–515.4942
6: $E(4, \lambda_1)$ and $C(\lambda_2, \lambda_3, p_c)$	$\alpha_0 = 2.7463$; $\alpha_1 = 0.5747$; $\lambda_1 = 0.0097$; $\lambda_2 = 0.0097$; $\lambda_3 = 0.0004$; $p_c = 0.7545$;	–515.0856
7: $E(5, \lambda_1)$ and $C(\lambda_2, \lambda_3, p_c)$	**$\alpha_0 = 2.6863$; $\alpha_1 = 0.5866$;** **$\lambda_1 = 0.0150$; $\lambda_2 = 0.0072$;** **$\lambda_3 = 0.0004$; $p_c = 0.7700$;**	**–514.8846**



(a) Comparison of empirical Cdf's and the Cdf's predicted by the EC-based ALT model

(b) Cox-Snell residual plot

Fig. 11.6 Illustration of statistical fittings of the resulting model

exponential distribution with mean of one (the straight line with slope of one going through the origin). Figure 11.7 shows the predicted reliability function, pdf and hazard rate of the lamp under the normal operating condition. The corresponding 95% bootstrap confidence interval for reliability function is also presented. The mean-time-to-failure can be easily obtained as 2397.2 h by setting $l = 1$ in Eq. (11.13).

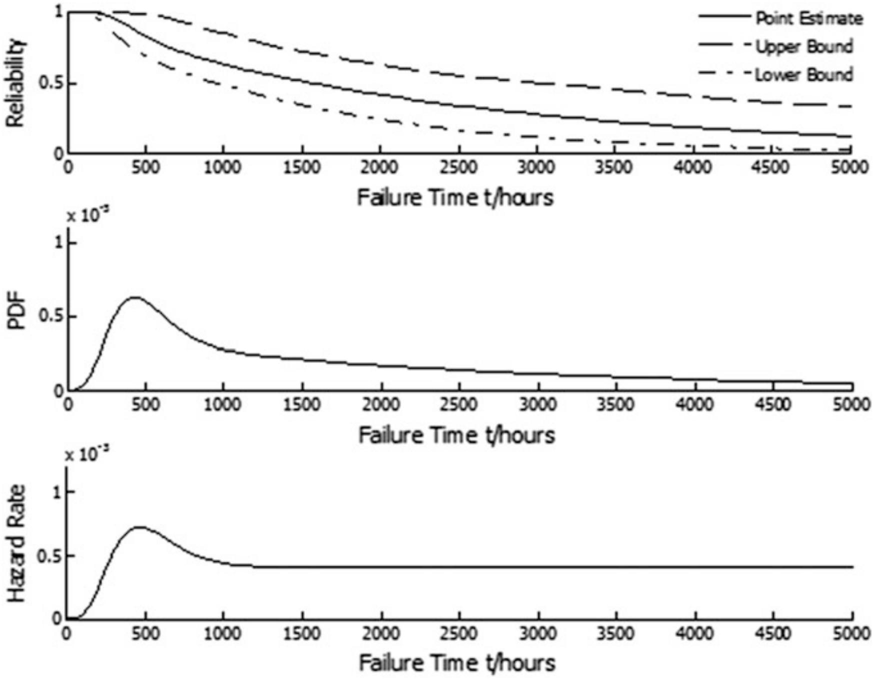


Fig. 11.7 Predicted reliability function with 95% confidence intervals, pdf, and hazard rate under 2 V using the resulting EC-based ALT model

11.6 Concluding Remarks

This chapter introduces a generic method for modeling ALT data using EC distributions, which belong to an important subset of PH distributions. Without assuming other particular probability distributions for failure times, such as extreme-value distributions, lognormal distribution, and gamma distributions, this method leads to an EC-based ALT model which can well represent the underlying failure time distribution that may be difficult to verify or even unknown. To automatically determine the model structure, both a mathematical programming approach and an MLE approach are developed for adaptively determining the number of phases and estimating the model parameters. The numerical examples demonstrate that the proposed generic method indeed provides practitioners, particularly in the area of microelectronics, with a convenient statistical tool for modeling ALT data. To the best of our knowledge, this is the first attempt to demonstrate the potential of using PH distributions in ALT data analysis.

Acknowledgements This work is supported in part by the U.S. National Science Foundation under grants CMMI-1238301 and CMMI-1635379.

References

1. Altiok T (1985) On the phase-type approximations of general distributions. IIE Trans 17(2): 110–116
2. Asmussen S, Nerman O, Olsson M (1996) Fitting phase-type distributions via the EM algorithm. Scand J Stat 23(4):419–441
3. Bagdonavičius V, Nikulin M (2002) Accelerated life models—modeling and statistical analysis. Chapman & Hall/CRC, New York
4. Bai DS, Chung SW (1992) Optimal design of partially accelerated life tests for the exponential distribution under type-I censoring. IEEE Trans Reliab 41(3):400–406
5. Bai DS, Kim MS, Lee SH (1989) Optimum simple step-stress accelerated life tests with censoring. IEEE Trans Reliab 38(5):528–532
6. Bhattacharyya GK, Soejoeti Z (1989) A tampered failure rate model for step-stress accelerated life test. Commun Stat Theory Methods 18(5):1627–1643
7. Doksum KA, Hóyland A (1992) Models for variable-stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution. Technometrics 34(1):74–82
8. Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap. Chapman and Hall, New York
9. Elsayed EA (1996) Reliability engineering. Addison Wesley Longman, New York
10. Elsayed EA (2012) Overview of reliability testing. IEEE Trans Reliab 61(2):282–291
11. Elsayed EA, Liao HT, Wang XD (2006) An extended linear hazard regression model with application to time-dependent-dielectric-breakdown of thermal oxides. IIE Trans 38(4): 329–340
12. Escobar LA, Meeker WQ (2006) A review of accelerated test models. Stat Sci 21(4):552–577
13. He Q, Chen W, Pan J, Qian P (2012) Improved step stress accelerated life testing method for electronic product. Microelectron Reliab 52(11):2773–2780
14. Johnson MA, Taaffe MR (1989) Matching moments to phase distributions: mixtures of Erlang distributions of common order. Commun Stat Stoch Models 5(4):711–743
15. Kielpinski TJ, Nelson W (1975) Optimum censored accelerated life tests for normal and lognormal life distributions. IEEE Trans Reliab R-24(5): 310–320
16. Lawless JF, Singhal K (1980) Analysis of data from life-test experiments under an exponential model. Naval Res Logist Q 27(2):323–334
17. Liao HT, Elsayed EA (2010) Equivalent accelerated life testing plans for log- location-scale distributions. Naval Res Logist 57(5):472–488
18. Marshall AH, Zenga M (2009) Simulating Coxian phase-type distributions for patient survival. Int Trans Oper Res 16(2):213–226
19. Marshall AH, McClean SI (2004) Using Coxian phase-type distributions to identify patient characteristics for duration of stay in hospital. Heal Care Manag Sci 7(4):285–289
20. Meeker WQ (1984) A comparison of accelerated life test plans for Weibull and Lognormal distributions and type I censoring. Technometrics 26(2):157–171
21. Meeker WQ, Escobar LA (1998) Statistical methods for reliability data. Wiley, New York
22. Nelson WB (1980) Accelerated life testing—step-stress models and data analyses. IEEE Trans Reliab R-29(2): 103–108
23. Nelson WB (1990) Accelerated testing—statistical models, test plans, and data analysis. Wiley, Hoboken
24. Nelson R (1995) Probability, stochastic processes, and queueing theory—the mathematics of computer performance modeling. Springer-Verlag, New York
25. Ni Z (2005) Moment-method estimation based on censored sample. J Syst Sci Complex 18(2):254–264
26. Olsson M (1996) Estimation of phase-type distributions from censored data. Scand J Stat 23(4):443–460

27. Onar A, Padgett WJ (2000) Inverse gaussian accelerated test models based on cumulative damage. J Stat Comput Simul 66(3):233–247
28. Osogami T, Harchol-Balter M (2006) Closed form solutions for mapping general distributions to quasi-minimal PH distributions. Perform Eval 63(6):524–552
29. Perez-Ocon R, Montoro-Cazorla D (2004) Transient analysis of a repairable system, using phase-type distributions and geometric processes. IEEE Trans Reliab 53(2):185–192
30. Riska A, Diev V, Smirni E (2004) An EM-based technique for approximating long-tailed data sets with PH distributions. Perform Eval 55(1–2):147–164
31. Ruiz-Castro JE, Fernandez-Villodre G, Perez-Ocon R (2009) Discrete repairable systems with external and internal failures under phase-type distributions. IEEE Trans Reliab 58(1):41–52
32. Sericola B (1994) Interval-availability distribution of 2-state systems with exponential failures and phase-type repairs. IEEE Trans Reliab 43(2):335–343
33. Sethuraman J, Singpurwalla ND (1982) Testing of hypotheses for distributions in accelerated life tests. J Am Stat Assoc 77(377):204–208
34. Tang LC, Sun YS, Goh TN, Ong HL (1996) Analysis of step-stress accelerated-life-test data: a new approach. IEEE Trans Reliab 45(1):69–74
35. Van Der Heijden MC (1988) On the three-moment approximation of a general distribution by a Coxian distribution. Probab Eng Inf Sci 2(2):257–261
36. Xiong C (1998) Inferences on a simple step-stress model with type-II censored exponential data. IEEE Trans Reliab 47(2):142–146
37. Yang G (2007) Life Cycle Reliability Engineering. Wiley, Hoboken
38. Zhang W, Liu S, Sun B, Liu Y, Pecht M (2015) A cloud model-based method for the analysis of accelerated life test data. Microelectron Reliab 55(1):123–128

Author Biographies

Haitao Liao is a Professor and Hefley Endowed Chair Professor of Logistics and Entrepreneurship in the Department of Industrial Engineering at University of Arkansas—Fayetteville. He received a Ph.D. degree in Industrial and Systems Engineering from Rutgers University in 2004. He also earned M.S. degrees in Industrial Engineering and Statistics from Rutgers University, and a B.S. degree in Electrical Engineering from Beijing Institute of Technology. His research interests include: (1) reliability models, (2) maintenance and service logistics, (3) prognostics, (4) probabilistic risk assessment, and (5) analytics of sensor data. He currently serves as Associate Editors for Journal of Quality Technology and IEEE Transactions on Reliability. He also served as Chair of INFORMS Quality, Statistics and Reliability (QSR) Section, and President of IIE QCRE Division. He is a recipient of the National Science Foundation CAREER Award in 2010, the winner of the IIE 2010 & 2013 William A.J. Golomski Award, 2013 Quality Control and Reliability Engineering (QCRE) Track Best Paper Award, and 2015 Stan Ofsthun Best Paper Award.

Ye Zhang received his Ph.D. degree (2014) in Systems and Industrial Engineering from The University of Arizona. He also received his B.S. degree (2008) in Industrial Engineering from Southeast University, and M.S. in Industrial Engineering (2011) from The University of Arizona. His research focuses on reliability models, and degradation data analysis and test plans.

Huairui (Harry) Guo is a Reliability and Statistics Specialist in the automotive industry. He obtained his PhD in Systems and Industrial Engineering from University of Arizona in 2004. He worked for ReliaSoft for over 10 years as the Director of the theoretical development group. Currently, he is working on reliability growth, warranty data analysis, degradation and accelerated life testing, and statistical method applications on quality and reliability engineering. He also provides internal trainings for quality and reliability engineers in the company. He is a certified quality engineer and a certified reliability engineer from ASQ, and a certified reliability professional. He is a member of IIE, SRE and ASQ.

Part IV
Component Degradation and Design in Accelerated Life Testing Data

Chapter 12
Thermographic NDT for Through-Life Inspection of High Value Components

Sri Addepalli, Yifan Zhao and Lawrence Tinsley

Abstract Non-destructive testing (NDT) are techniques used to detect and characterise flaws that occur in materials from manufacture through to evaluating the health of the material without causing further damage to the component. With the development of cutting-edge technology over the last two decades, active thermographic NDT has grown considerably as a field. Access to lower cost, more portable hardware with higher performance has fuelled a major drive in research to develop analytical techniques and widen the applicability of thermography in order to exploit its advantages as a low-cost and non-contact inspection technique. While passive thermography is a heavily standardised process, active thermography is considerably lacking in industrial standards. Development of these standards represents an opportunity for research in the field of active thermography to be a part of that process. Recent industry pressure regarding research in NDT has developed a demand for NDT techniques to be quantifiable, and linked directly to material properties, thus allowing an estimation of remaining useful life (RUL) in order to maximise product value. There have been significant research developments in the field over recent years. Through developments in signal processing of thermography data, inspections would enable repeatable, quantifiable benchmarking of samples, which would allow automation of carefully controlled quality checks; and the measurement of thermal properties of the material, which would allow estimation of components' RUL. Addressing these challenges would increase the deployment of thermography in industry, enhance the toolset of through-life engineering, and significantly improve the competitiveness of industries which embrace these developments.

Keywords Pulsed thermography · Non-destructive testing · In-service damage · Automated NDT

S. Addepalli (✉) · Y. Zhao · L. Tinsley
EPSRC Centre for Through-Life Engineering Services, Cranfield University,
Cranfield, UK
e-mail: p.n.addepalli@cranfield.ac.uk

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_12

12.1 Introduction

The 20th century has seen a rapid increase in the production of smart and innovative systems. Thanks to the developments in the area of applied computing which has revolutionised design of systems right at the component level. Age old techniques have now been replaced with a modern outlook to bring in the best performance that particular system can provide. With the introduction of such next generation parts comes the responsibility of maintaining them so that their life can be increased. This also means that, with the refinement in the design and manufacture of such complex engineered parts, traditional maintenance techniques need modification to help maintain the integrity of such systems thereby extending the life of the component and preventing any premature failure during service. However, with smart and advanced systems, implementing advanced maintenance techniques is the way forward.

Non-destructive testing or NDT has a major role to play in the maintenance activities of manufacturing and maintenance industries as they help assess the health of the component there by advising any additional repair that may be required by the component and or the system. Due to its nature of not altering the structural integrity of the component, and ability to determine the health of the component remotely using scientific methods, NDT has found its way into the industrial sector making huge maintenance savings. With the introduction of advanced innovative materials and products, it is a challenge for age old traditional NDT techniques to be able to provide concrete information about the health of the system. For example, detecting disbonds and monitoring their growth especially in graphite based composites might be a challenge for traditional NDT techniques where advanced techniques such as thermography and immersion pulse-echo ultrasonic technique might be the only way forward [1].

With the introduction of next generation systems that are far more advanced and superior than the existing counterparts, the industry is forced to look for supporting maintenance systems that could help cater the service needs those complex engineered systems might need in the future. With complete shift in service strategies where the service burden has fallen back to the manufacturers [2], high value manufacturing (HVM) industries are turning to cheaper and intelligent automated systems that could provide fast and reliable service critical decisions that reduces service lead times and improves confidence in understanding the health of the system during maintenance.

The current focus for HVM based industries is to develop smart and innovative maintenance solutions for their advanced parts that are currently in the research and development stage so that the maintenance technologies mature with time and are ready for use when such parts get into mainstream market. It should be understood that with the advent of advanced manufacturing systems such as additive layer manufacturing, there is a huge challenge to understand the behaviour of such parts in harsh environments and hence it is of utmost importance to have maintenance technologies in place that could investigate the health of parts during service.

Furthermore, it is key to retain that tacit knowledge and convert the maintenance activities into an intelligent decision making system whereby repair strategy decisions could be automated through the access of knowledge based database systems, thus performing the entire maintenance activity autonomously with a high level of reliability. This chapter mainly presents an alternate automated inspection approach where the entire activity right from inspection to analysis is fully automated.

12.2 Automation Approach

Within the area of material degradation assessment, there has been constant development in automating the data capture with evidence especially in the HVM context. Literature does show the use of advanced signal and image processing techniques and key challenges to integrate technologies that use temperature measurements and relate them with material properties [1, 3, 4]. There are many advantages in the use of pulsed thermography, an active thermography technique, where a surface temperature decay curve response due to flash heating brings in sub-surface anomalies as surface temperature colour maps there by providing enhanced structural information of the part being inspected in a non-destructive manner [3–7]. Together with automation, this NDT tool provides a much faster, repeatable and reliable inspection process leading to reduction in repair and other maintenance strategies for such HVM parts.

Figure 12.1 below is a process flow diagram explaining the automation approach that has been successfully implemented as part of the research in the area of HVM component degradation analysis. The automated inspection process which starts with the operator starting an industrial robotic arm to pick the part; in this case the robot follows a pre-defined tool path. Once the part is picked up, the robot arm presents it to the pulsed active thermography system ready for inspection. As soon as the part arrives to the predefined inspection position, the inspection system triggers an optical pulse flash followed by the infrared camera (part of the inspection system) capturing the surface temperature decay profile of the part. The surface temperature data now goes through a post-processing stage, where it gets reconstructed using the thermographic signal reconstruction or TSR method [8–10]. The data then gets transferred and saved on to the cloud. The in-house analysis

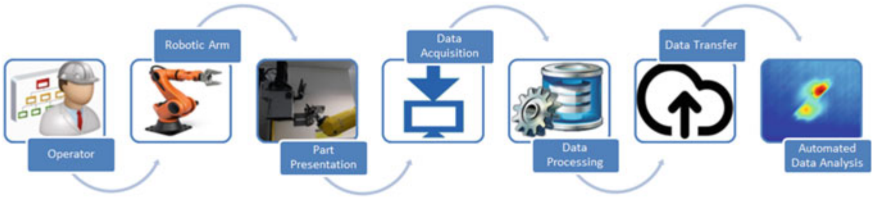


Fig. 12.1 An automated pulsed thermography inspection system

software, then retrieves the post-processed data from the cloud, and sifts through the data and presents it in the form of infrared images or thermograms and temperature time plots. Thus the entire system in effect becomes a one click system where the entire process right from part presentation to the inspection system through to the final analysis is fully automated. The main highlight of the system is its ability to autonomously perform the inspection to a user defined process, where the sentencing of the part can be achieved by easily providing a temperature threshold which increases the signal-to-noise ratio or SNR of the system which then clearly identifies the damage area. This technique benefits from not just the multiple visualisation techniques employed to represent the acquired data, but also the adaptability of the inspection system to be mounted on to the robotic arm thereby building the systems capability to inspect large components. Another additional feature of this method is the ability to present a complex geometry part in pre-defined multiple angles maintaining repeatability of part presentation and to cover maximum surface area especially around curvatures.

12.3 Multiple Data Visualisation Techniques

For demonstrating the strength of the inspection technique and its ability to be represented in multiple visualisation techniques, carbon fibre reinforced polymers or CFRP laminates having varying levels of controlled impact damage was chosen. The following sub-sections describe the laminate, the impact test and the pulsed thermography inspection results.

12.3.1 The Impact Test

CFRP laminates made of unidirectional fibres with Hexcel M21 resin were made through autoclave process and were cut to samples of 150 × 100 × 4 mm. These laminates were then subjected to a drop test using the instron drop test machine for pre-defined drop energy levels between 5 and 30 J. The drop hammer had a hemispherical indentor of 16 mm diameter and weighed 2.281 kg. Energy levels were calculated based on the drop height where the energy equivalent (E) can be calculated using the equation;

$$E = m \times g \times h \tag{12.1}$$

where, m—mass of the indentor = 2.281 kg; g—acceleration due to gravity = 9.81 m/s^2 and h—the drop height in meters.

For the purpose of this study only three energy levels, 10, 20 and 30 J were selected to present the results. The full detailed information of the work and the complete set of results can be found from Zhao et al.'s paper on coefficient clustering analysis for composite laminate impact damage assessment [11].

12.3.2 Pulsed Thermography Results

The drop test machine produced visible damage on the laminates as seen in Fig. 12.2 below. It was noticed that the 10 J impact did not show any direct surface breaking damage other than a physical deformation, however the 20 and 30 J laminates showed associated surface cracking which is dependent on the drop test energy level. To understand if the damage could be detected in the first place, the laminate was subject to pulsed thermography inspection. It should be noted that the inspection was performed from the back surface and not the impact side. The pulsed thermography system comprised of 2 xenon flash lamps enclosed in a reflective hood and powered by a capacitor bank system, a computer control unit and a FLIR SC7600 indium antimonide (InSb) sensor based infrared radiometer operating at wavelengths 3–5.1 µm.

Pulsed thermography could be described as the technique that observes the surface temperature decay profile when excited with an instantaneous heat pulse. It has been demonstrated that when a homogeneous material cools, the heat diffusion through the material which is dependent on its thermal conductivity is constant and thus the surface cools evenly. As soon as a discontinuity occurs along the heat diffusion path, there is localised change in diffusion which appears as either a hot or a cold spot on the surface temperature. Thus the surface temperature at a time (T(t)) can be determined using the following equation [12];

$$T(t) = \frac{Q}{\sqrt{\pi \rho c k t}} \left[1 + 2 \sum_{n=1}^{\propto} R^n \exp\left(- \frac{n^2 L^2}{\alpha t} \right) \right] \tag{2}$$

where Q—heat pulse energy, ρ—material density, c—materials specific heat capacity, k—thermal conductivity of the material, L—defect depth, α—thermal diffusivity and R—the air gap interface's coefficient of thermal reflection. This equation describes the heat flow through a semi-infinite homogeneous plate where the heat decay characteristics are governed by the plate's material property. It has also been established that when a time temperature plot is produced in the logarithmic domain, the slope for a standard area occurs with a slope of −0.5 and for that of the damage will deviate from the this value [13].

The result obtained from a 30J impact sample (see Fig. 12.3) shows that the damage created due to the drop test is clearly visible with the highest damage being represented by the red area in the image. As mentioned above, the curve representing the sound area has a steady cooling characteristic over time with damage

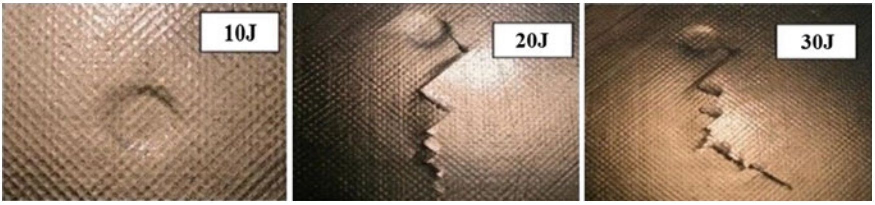


Fig. 12.2 Individual laminates showing the impact area (adapted from [11])

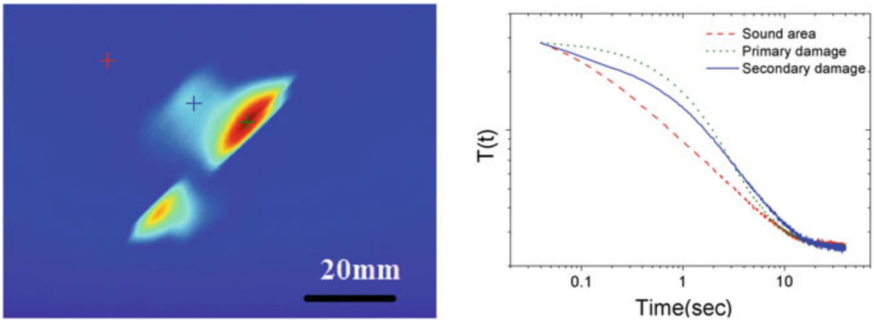


Fig. 12.3 RAW Thermal data from 30 J impact sample with the thermogram on the *left* and a representative temperature time plot on the *right* (The *coloured curves* in the plot are data points obtained from pixel locations represented by *colour markers* from the image on the *left*) (adapted from [11])

areas showing distinct deviation from the curve obtained from the sound area as established in the literature [14–16]. It was also noticed that the maximum damage in this case was around the impact area and not directly at the indentor contact point as seen from the visual image above (see Fig. 12.2).

12.3.3 Damage Area Measurement

One of the key areas of algorithm development was to present the ability of measuring the damage area where visualisation in the binary form is used to measure the damage shape. The following visualisation was created from laminates that have undergone varying energy impact test.

From the binary maps (see Fig. 12.4) the damage area was estimated to be 154.92, 648.84 and 1283.04 mm^2 for the 10, 20 and 30 J laminates respectively. Further, reconstructed visualisation produced the images showing varying damage levels created during impact (see Fig. 12.5).

Fig. 12.4 Binary visualisation for impact tested laminates

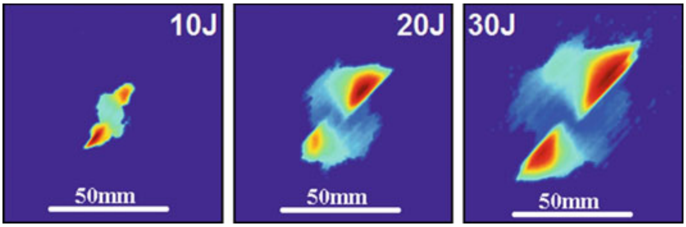


Fig. 12.5 Damage visualisation of the reconstructed data from the coefficient clustering analysis (CCA) method

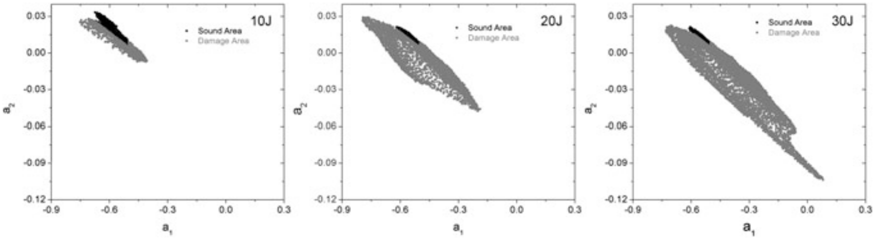


Fig. 12.6 Scatter plot obtained from the first and second order coefficients for all impacted laminates

12.3.4 Coefficient Cluster Analysis (CCA)

The following is the data obtained from the coefficient cluster analysis or CCA method where the scatter below is obtained from selecting the regions around the impact area for a size of 200 × 200 pixels.

The principle of the CCA method is to use a low order polynomial model to fit the raw temperature decay curve initially, and then a clustering method is applied into the scatter between the identified parameters to classify the pixels from sound and damage areas. As shown in Fig. 12.6, the x and y axes denote the coefficients of the first and second order terms. The grey dots are classified as the pixels from the damage area and the black points are the pixels from the sound area.

The results from the clustering can then be reversed back to the space domain as demonstrated in Fig. 12.4.

The results presented in this section are from a case-study where CFRP laminates were subjected to impact damage and inspected using pulsed thermography. It was found that the information obtained from the inspection could be presented in the form of thermograms and temperature time plots for selected pixels, which has been well established by various studies [8, 10, 16]. In addition to the established techniques, this paper presents additional visualisation techniques where damage in a localised area could be plotted as against single pixel information [11]. Further the reconstructed data showed the overall damage area and provided a much better differentiation between primary and secondary damage when compared with the data representing sound area. It was also established that in this case, the damage area continued to grow with increasing impact energy. This was also visualised from the scatter plot (see Fig. 12.6) where the scatter for the damage area continued to increase with impact energy.

12.4 Summary

This chapter presents the developments in the area of automated thermographic detection by providing an overview of the entire automation process right from picking the part all the way to the post analysis and the associated tool sets that will help the operator make those critical decisions as part of the inspection process. It has been demonstrated that, in addition to the strength of the inspection technique, there is continuous development of tool sets that help visualise data in a way that can be easily interpreted. All the visualisation tool sets were developed in-house and the results were compared with standard commercial systems that can validate the original inspection data. It is the aim that these developments will enhance not just the decision making capabilities but also provide a higher confidence level when it comes to sentencing the part especially during maintenance through automation.

Acknowledgements The authors would like to thank Professor Peter Foote from Cranfield University for providing the CFRP specimens. This work was supported by the EPSRC Centre for Innovative Manufacturing in Through-life Engineering Services (Grant number EP/I033246/1).

References

1. Genest M, Martinez M, Mrad N, Renaud G, Fahr A (2009) Pulsed thermography for non-destructive evaluation and damage growth monitoring of bonded repairs. Compos Struct 88(1):112–120
2. Meier H, Volker O, Funke B (2011) Industrial Product-Service Systems (IPS^2)—paradigm shift by mutually determined products and services. Int J Adv Manuf Technol 52:1175–1191

3. Mehnen J, Tinsley L, Roy R (2014) Automated in-service damage identification. CIRP Ann Manuf Technol 1
4. Addepalli S, Tinsley L (2015) Active thermography in through-life engineering. In: Redding L, Roy R (eds) Through-life engineering services, decision engineering. Springer International Publishing, pp 117–127
5. Maldague X (1999) Infrared thermography and nondestructive evaluation, May 1999 [Online]. http://www.ndt.net/article/v04n05/irtnde/irtnde1.htm. Accessed 26 Jan 2014
6. Ibarra-Castanedo C, Piau J, Guilbert S, Avdelidis N, Genest M, Bendada A, Maldague XPV (2009) Comparative study of active thermography techniques for the nondestructive evaluation of honeycomb structures. Res Nondestr Eval 20(1):1–31
7. Maldague XPV (2002) Introduction to NDT by active infrared thermography. Mater Eval 60(9):1060–1073
8. Shepard SM (2007) Flash thermography of aerospace composites. In: IV Conferencia Panamericana de END, Buenos Aires
9. Balageas DL, Roche J-M, Leroy F-H, Liu W-M, Gorbach AM (2015) The thermographic signal reconstruction method: a powerful tool for the enhancement of transient thermographic images. Biocybern Biomed Eng 35(1):1–9
10. Shepard SM (2003) Temporal noise reduction, compression and analysis of thermographic image data sequences. United States of America Patent US Patent 6516084 B2
11. Zhao Y, Tinsley L, Addepalli S, Mehnen J, Roy R (2016) A coefficient clustering analysis for damage assessment of composites based on pulsed thermographic inspection. NDT&E Int 83:59–67
12. Lau S, Almond D, Milne J (1991) A quantitative analysis of pulsed video thermography. NDT&E Int 24(4):195–202
13. Shepard SM, Hou J, Lhota JR, Golden JM (2007) Automated processing of thermographic derivatives for quality assurance. Opt Eng 46(5):051008
14. Shepard SM (2011) Advances in pulsed thermography. Proc SPIE 4360:511–515
15. Vavilov V, Maldague X, Dufort B, Robitaille F, Picard J (1993) Thermal nondestructive testing of carbon epoxy composites: detailed analysis and data processing. NDT&E Int 26(2):85–95
16. Ball R, Almond D (1998) The detection and measurement of impact damage in thick carbon fiber reinforced laminates by transient thermography. NDT&E Int 31(3):165–173

Chapter 13
Engineering Support Systems for Industrial Machines and Plants

Youichi Nonaka, Takahiro Nakano, Kenji Ohya, Atsuko Enomoto, Gábor Erdős, Gergely Horváth and József Váncza

Abstract In the business of industrial machines and plants, rapid and detailed estimates for planning installation, replacement of equipment, or maintenance work are key requirements for meeting the demands for greater reliability, lower costs and for maintaining safe and secure operation. These demands have been addressed by developing technology driven by IT. When replacing equipment at complex building or plants with high equipment density, the existing state of the installation locations and transportation routes for old and new equipment need to be properly measured. We have met this need by developing parts recognition technology based on 3D measurement, and by developing high-speed calculation technology of optimal routes for installation parts. This chapter provides an overview of these development projects with some real business application results.

13.1 Introduction

Satisfying the complex web of client requirements and site conditions when constructing a new facility such as elevator system, power plant, chemical plant or oil refinery requires a wide variety of engineering work. This includes environmental assessment, civil engineering/construction, equipment design, equipment procurement, installation, and trial operation/handover. Those maintenance also requires advanced and detailed engineering work to diagnose component equipment, machinery, and devices, and to repair/replace them as needed to maintain safe and stable operation. When providing engineering services to clients for new plant construction or maintenance, detailed and rapid estimates of costs and work schedules need to be created. To meet these needs, we have developed several

---

Y. Nonaka (✉) · T. Nakano · K. Ohya · A. Enomoto
Research & Development Group, Hitachi, Ltd., Yokohama, Japan
e-mail: youichi.nonaka.ym@hitachi.com

G. Erdős · G. Horváth · J. Váncza
Institute for Computer Science and Control, Hungarian Academy of Sciences,
Budapest, Hungary

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_13

technologies designed to enable more advanced engineering work through use of latest information technology (IT).

This chapter looks at retrofit and replacement of industrial machines and plants. These facilities of between 20 and 50 years old sometimes only have original design drawings in two dimensions (2D), or have undergone so much maintenance over the years that their cables, plumbing or equipment systems have become unrecognizable from the original drawings. Basing a project plan on the existing state is an essential requirement in these cases.

Retrofit/replacement projects consist mainly of removing the items that need to be upgraded, and installing the new equipment. When planning each operation, a key requirement for large plants is to identify the actual state of the surrounding environment to answer questions such as whether there are any obstacles in the transportation routes for removal and installation, or whether reliable connections can be made to existing equipment items. Specifically, the site is surveyed to identify locations to be added or moved, locations that have been transformed by many years of operation, and other site-specific issues. The results of these studies are then used to create structural designs and process designs.

To increase the efficiency of processes ranging from design to installation, and to eliminate the need for skill in site studies, we have worked on using three dimensional (3D) measurement by long-distance contact-free laser scanners to enable rapid measurement of existing site conditions. 3D laser scanners have recently come into wide use in fields such as civil engineering, construction, and surveying. But measured point clouds contain noise, and since point cloud data is massive, generating as-built models requires extensive manual labour.

Also, when retrofitting or replacing substation equipment, a large amount of plant-assembled and plant-inspected equipment items are successively installed at the site. The equipment is heavy, so they are lifted by crane for transport, positioning, and connection work. These processes require studying the installation sequence and creating work plans after taking into account difficulties in making equipment parts fit each other in three dimensions, and the temporary placement of equipment delivered to the site. That is, the process of planning the installation of substation equipment items requires a lot of experience and specialized knowledge.

We met these requirements by developing an “as-built” modelling technology based on 3D measurement and a technology for planning approaches to retrofit/replacement based on 3D models.

13.2 Related Works and Technology Trends

In industrial engineering (IE) there is a common understanding that the primary prerequisite of performing maintenance, repair and overhaul (MRO) activities is to have a reliable digital **semantic model** of the overall target object [11, 13, 27]. Such a model, which is referred to in the closely rated field of architecture, engineering and construction (AEC) as **building information model** (BIM) represents the complex industrial facility in terms of its components, along with their geometric and any other relevant properties and relationships. Unlike a traditional computer-aided design (CAD) model it is a semantically rich representation that can be used by a number of different stakeholders of a facility with relatively long life-cycle, including MRO planners and operators, too [18, 28]. However, just such industrial objects—like elevators, plants, thermal or nuclear power stations, manufacturing facilities to name a few—have in many cases only two dimensional blue print documentation (if any) [19], and it is rather the rule than the exception that there are mismatches between any kind of model and the reality [9]. Planners responsible for MRO activities have to face risks due to errors of unrecorded modifications, deformations, as well as missing records of incidental equipment such as suspending fixtures and cranes. When retrofitting or de-constructing complex industrial objects, a correct model of the work environment can facilitate both safety and efficiency. Tailoring the model to reality time and again is also essential when monitoring the progress of construction projects and registering what has (or has not) been built according to plan or specifications [26]. Hence, it is not enough to create a semantically rich model of the industrial object, but this model should capture its actual **as-is state** [23].

In the past decades, these double requirements opened broad and overlapping research fields both in IE and AEC. With the advancement of computer vision and especially **3D laser scanning** technologies became possible to scan the surface of even very large-scale objects and to create their as-built representation in terms of a set of points with 3D Cartesian coordinates—a so-called **point cloud** [1, 3, 4, 12]. In parallel, methods for the efficient storage and retrieval of this often enormous amount of measurement data have been developed, too [24]. However, the **reverse engineering** problem of converting this raw representation into a concise, semantically rich model remained a research challenge till today [12, 28, 31].

A detailed survey of methods supporting the (semi-)automated **reconstruction of as-built building information models** out of laser-scanned point clouds is provided in [28]. Accordingly, as components of a complex object have shape, identity and relationships, the overall problem involves three essential sub-problems: geometric modelling, object recognition and object relationship modelling. While advance in all the above research areas have been achieved, and even some specialized commercial systems have appeared, so far there has been found no integrated, generic solution for building up automatically a structural model of a complex industrial object departing from its point cloud data.

The actual solutions vary with the domain; for instance, in AEC much effort was put into developing methods that can recognize **indoor scenes** [22] and the most characteristic components of a building, such as floors, walls, cabinets, ceilings, and openings like doors and windows [28, 31]. Recognition is typically concentrating on the surface of objects, via polygonal meshes and parametric surface models fitted to the point cloud. Hence, these methods generate models of complex objects in terms of structured surface meshes. A recent development aims at a reconstruction of buildings which maintains room **topology** and global wall connectivity [21]. Having a specific scope on construction sites, [29] tackles the recognition of heavy dynamic construction equipment like crawler cranes. In industrial engineering, the well-proven notion of **features** is applied almost unanimously, whose recognition can depart from a mesh-type input data that is fitted to the measurement point cloud [30]. A recent survey provides a comparative analysis and outlook of academic and state-of-the-art commercial methods capable of recognizing **as-built 3D layout** and in particular, **pipeline systems** in large-scale civil infrastructure facilities by making use of photo- and video- grammetry, as well as terrestrial laser scans [26]. Under practical conditions, due to complexity, noise level and incompleteness of data the methods are though hardly applicable without extensive (and expensive) human assistance. Hence, more research is expected, especially in a direct collaboration of academy and industry [26].

Our earlier research aimed at **adapting** the existing CAD model of a complex industrial object to the point cloud measured on its actual surface was just an instance of such a joint study [9]. The workflow included the efficient storage of massive measurement data, segmentation of a triangulated mesh-based CAD model into features, as well as matching and adapting the features to the data. The method was applied in a real-world setting, using the CAD model and point cloud data of an industrial plant that contained planar and cuboid objects, and complex and dense systems of pipes. Continuation of this research led to a method that was capable of reconstructing the **structural model** of as-built industrial facilities purely from on-site point cloud measurement data [10]. Focus was set on finding the internal structure of complex objects hidden behind the massive point cloud by exploiting connectivity information in the data and the linear characteristics of the typical components such as pipes, beams or other structural elements. This novel method, along with examples of its application in a specific domain will be presented in some detail in Sect. 13.3 below.

The other main field of research related to our topic is automated **disassembly and assembly planning** which involves in the MRO domain two kinds of sub-problems: (dis)assembly sequence planning, as well as (dis)assembly path planning. Sequence planning concerns the problem of finding a feasible sequence of operations that remove (put) components of a complex object from (to) their place. While the reconstruction of the object's structure is such a prerequisite of sequence planning which can be resolved in specific cases (see above, or [1] in particular), so far there are only few and scattered attempts to derive the sequence of (dis)assembly operations from the model of the object at hand [14]. However, in the past two decades much effort was put into the solution of the problem of transporting components, equipment and other objects in a relatively densely occupied industrial environment [6]. This latter so-called carry-in and carry-out problem (whose generic version is the classic piano movers' problem) is essentially a task for **path planning** which centers around generating a collision-free path from an initial picking point to a target point in an environment filled with obstacles. Further to accounting for part and obstacle geometries, as well as for physical forces such as gravity or friction, the path has to be optimal according to some criterion like energy, time, safety, etc. [19]. A recent review provides a comprehensive taxonomy and characterization of (dis)assembly path planning problems, together with a well-structured presentation of up-to-date solution approaches [14]. A generic, broadly applied method of path planning in constrained spaces is that of **road mapping**. Even though depending on its actual application domain like robotics, (dis)assembly, unmanned aerial vehicle (UAV) planning, MRO planning, etc., road mapping developed numerous variants, its core concept is based on a network of collision-free configurations where adjacent nodes can simply be reached from each other. Road maps can be pre-computed by probabilistically sampling the space [16] before searching for a solution (by using some classical shortest-path graph search or A* algorithm), or generated on the fly, as in the case of rapidly-exploring random trees. In any case, finding narrow passages between collision-free areas poses a serious challenge, especially for planners operating in a 3D cluttered environment. The key to that issue is the characterization of the space and adapting appropriately the sampling strategy which heavily relies on collision detection. For instance, so as to increase computational efficiency we applied an octree-based voxel representation of the free space, combined with its parallelized exploration [8, 19] (for details, see Sect. 4 below).

Finally, carry-in and carry-out operations are executed typically by **cranes**. While planning the path of such auxiliary equipment is in most of the cases out of the scope of investigations [14], the practically highly relevant problem of mobile crane walking and path planning is tackled in [17]. Here, crane configurations are considered together with typical site constraints and the geometry of lifted equipment. The proposed method determines the pick and (collision-free) operation areas, and then calculates the walking path of the crane. A related problem is **crane lifting** in a complex environment when a collision-free and cost-optimal lifting path is to be generated by considering inputs such as the plant environment, crane mechanical data, crane position, as well as pick and end lifting configurations. An overview of state-of-the-art computer-aided crane lift planning methods is presented in [5], along with a specific genetic algorithm based technique that proved to be highly efficient thanks to its parallelized implementation.

13.3 Recognition and Modelling Technology

13.3.1 The Object Recognition Problem

The object recognition and modelling technology is based on some generic assumptions that ensue from the application domain. First, even though measurement data may be acquired from a number of different positions of the scanner, all points are registered in the same reference coordinate system. The proprietary point cloud data format of the actual scanning system is also transformed to a uniform representation. Multiple scans can to some degree alleviate the difficulties caused by occlusion, but the measurement data remains intrinsically noisy and partial. As for the structure of the scanned object, one can assume that it is constructed out of typically linear extruded elementary components, such as pipes, beams, pillars, walls or cuboid objects. Some of these objects may be even of standard size (like various types of beams). However, we emphasize that the availability of the CAD model of the object is not required.

Hence, the **inputs** of the recognition process are the following:

- 3D point cloud of the measured complex object,
- prior knowledge of the types of its elementary components, and optionally,
- additional information on the exact geometries of the potential elements, such as catalogue of standard beams.

The **result** of recognition process is a compact representation of the measured object consisting of its identified elementary components, together with their actual geometric parameters and their connectivity relations. Furthermore, each point of the cloud has to be indexed either with the components found or marked as unidentified.

The strong engineering motivation of the application implies twofold performance **criteria**: minimizing overall processing (including manual and computational) time, and achieving as high recognition accuracy as possible.

13.3.2 Assumptions and Representation

The representation and method developed for transforming the large-scale 3D point cloud data into a structured model of a complex object is based on some fundamental engineering principles.

- **Aggregation** is applied when collecting points of the cloud into a discrete, uniformly sized 3D grid structure and working with these voxels instead of the points in some of the calculations.
- **Filtering** is used to remove noise from the data.
- **Segmentation** is applied to decompose a larger space investigated into regions of manageable size. The recognition process can run in each region simultaneously, while some overlap between the regions warrants that connectivity information is not lost.
- **Connectivity** of voxels is maintained and exploited so as to recognize topological relations between elementary components whose surface is represented —even partially—by the voxels.
- **Linearity** of elementary components that build up a complex engineering object is assumed.

Hence, the representation has altogether four layers. The basic representation is that of the registered measurement data points, given as coordinates in a common Cartesian system. Next, points are clustered into voxels that are signified by their centroid points. Then, neighboring voxels are captured in a **voxel connectivity graph** (VCG) where nodes denote voxels, while edges stand for any two voxels which are adjoining in space. A region under study is typically covered by a number of disjoint VCGs. Finally, each VCG has also a more refined model where the linearity of components is explicitly exploited. This is a so-called **branch connectivity graph** (BCG) where the nodes stand for typically linear branches composed of specific connected subsets of adjacent voxels of a VCG, while edges represent connections between the branches. The BCG provides a more articulated representation of the measurement data and hints at the presence of typical object types.

13.3.3 Workflow of Object Recognition from Point Cloud Data

Specific illustrations for the stages are taken from a case-study performed in an industrial domain, where recently the method has been applied routinely for elevator renewal.

As the case-study Fig. 13.1 indicates an image of aged-elevator shaft scanning work in the renewal business. 3D laser scanners on the pit and on cage get point cloud data refers shaft, rails, doors, cage, and so on, those are deformed and/or tilted by many years of operation. That is, the purpose of object recognition is to detect objects such as shaft, rails, doors, cage, and quantify those deformations and tilting automatically.

A workflow has been developed for solving the above problem that consists of the stages of **preprocessing**, **point cloud filtering and connectivity graph construction**, as well as elementary **object and connectivity recognition**. Figure 13.2 presents this workflow, while the next subsections describe in short the major processing stages (for more details, see [10]).

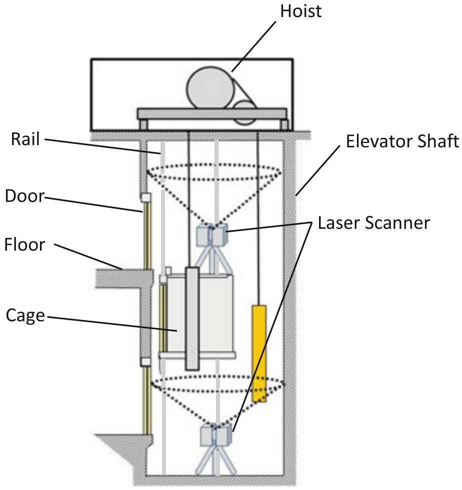


Fig. 13.1 3D laser scanners in elevator shaft (*vertical* cross section)

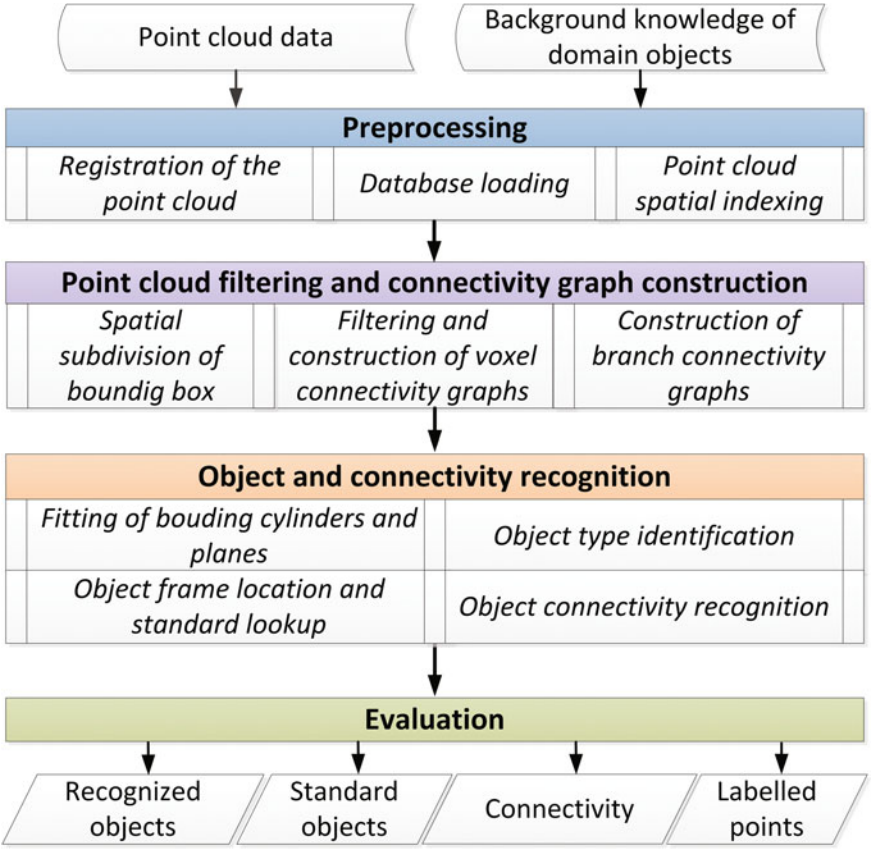


Fig. 13.2 Workflow of the object recognition process

13.3.3.1 Pre-processing

First, an affine transformation is performed to register points taken from various scanner locations in common reference coordinate system. Next, the space is decomposed into 3D **voxels** of a given size. However, efficient storing and querying large datasets containing up to even billions of points requires appropriate indexing schemes and database management techniques. Since the most frequent operation is bounding box query, such a spatial indexing is used that stores data indexed by the basic voxels so that points located close to each other in the actual domain are stored also physically close to each other in the database. The indexing scheme applies an octree-based decomposition of the space [24].

Figure 13.3 shows the raw point cloud data captured in the shaft of an elevator. This relatively small dataset contains ca. 40 million points.

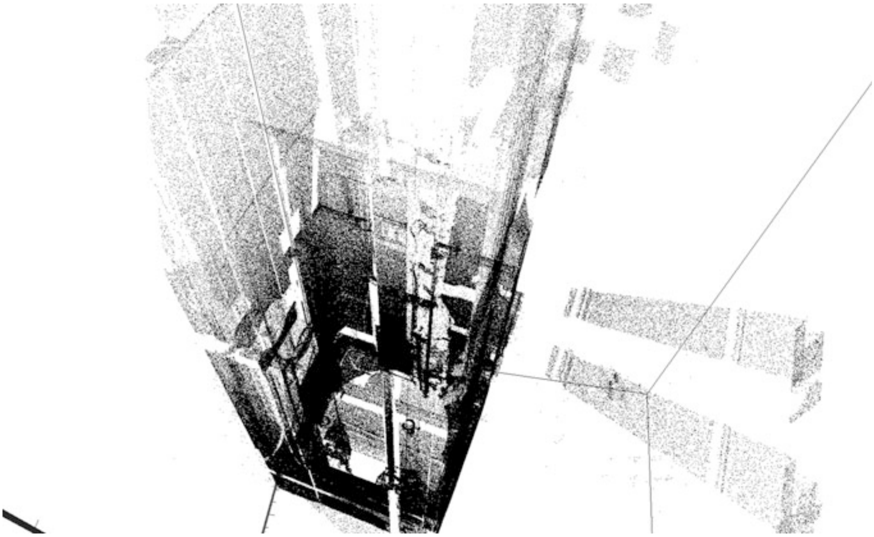


Fig. 13.3 Point cloud data of the elevator shaft

13.3.4 2-Point Cloud Filtering and Connectivity Graph Construction

In order to keep the size of the raw point data as well as the complexity of recognized connected object structures manageable, the space of the complete object is first **segmented** into disjoint spatial regions. Points belonging to the same region are processed together, while data of different regions are processed independently (and, optionally, in parallel). The size of a region depends on the number and granularity of the measured points. The generic rule is that data belonging to one region should fit into the memory of the actual computing facility. For instance, in the actual elevator case study there is no need to segment the space into regions.

The point cloud is collected through a series of on-site measurements, hence due to occlusion, shadowing, and inaccessibility on the one hand, and reflections on the other hand, the data is incomplete and burdened by noise. **Filtering** and **VCG composition** remove the noise from the input data and select such connected subsets of voxels that are good candidates for object recognition. The procedure composes a VCG where both the number of points in each voxel and the number of voxels in each connected component are over some specific, pre-determined thresholds. Voxels (and included measurement points) not meeting any of these criteria are discarded in the course of an iterative filtering process. By interleaving filtering and VCG construction, both scattered and isolated points are removed from

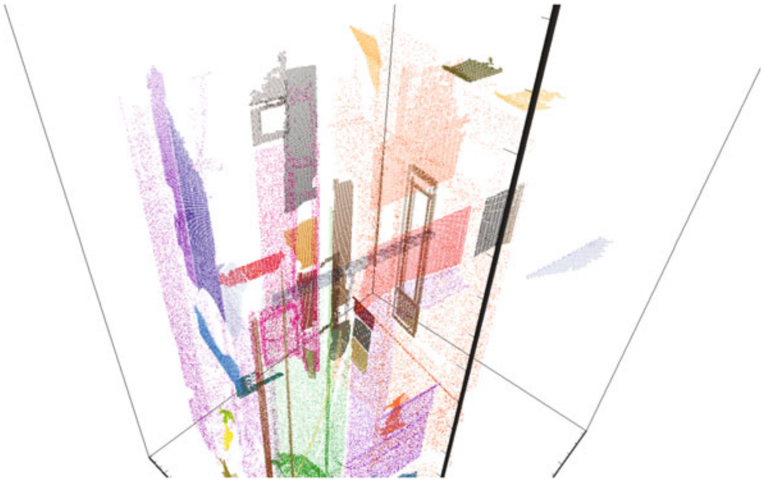


Fig. 13.4 VCGs of the elevator data

further processing. Figure 13.4 shows the VCGs generated for the elevator data: voxel size is 1 × 1 × 1 cm, minimum point density is 25 point/voxel, and at least 500 voxels should be connected. This way one gets ∼50 disjoint connected sets of ca. 500.000 voxels.

Subsequent steps of the workflow can focus on those areas of the space that are not only densely populated by points, but contain also candidates of large enough compound objects. However, in a VCG some disjoint connected components can be too complex and unstructured for recognition. Hence, these are deconstructed into smaller connected subsets by exploiting the linearity assumption. The typical elementary components are extruded objects like pipes, beams (with various profiles), rails, etc., which can be represented by connected **voxel branches** stretching in some characteristic direction. Planes are extreme types of such extruded linear objects. **The BCG construction** method applies projection and a specific region growing method to find both quasi-linear structures as well as their connections. First, in a given direction linear arrangements of connected voxels (so-called fibers) are sought with a length over a threshold. Fibers with adjacent voxels form a branch which is augmented with isolated voxels in its immediate proximity. Connectivity of branches that have adjacent voxel pairs is recorded. Finally, after removing voxels of the BCG found so far the procedure is iteratively repeated for other directions. Figure 13.5 provides two different looks of the BCGs generated for the elevator data (with minimal fibre length of 50 cm). Voxels of the same color belong to the same branch of the BCG.

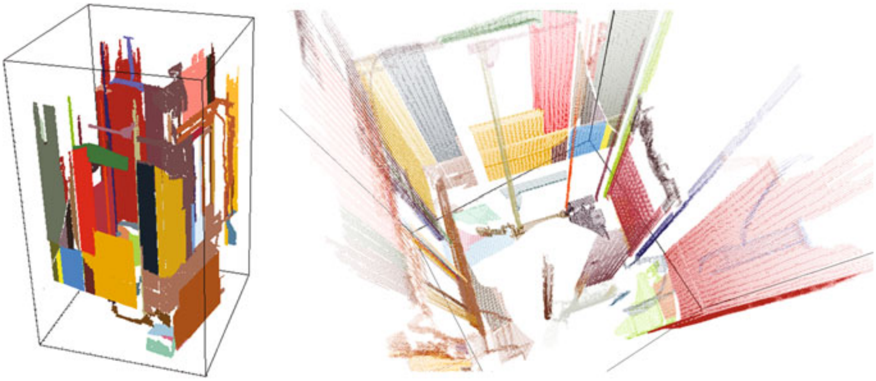


Fig. 13.5 BCGs of the elevator data depicted in two different views

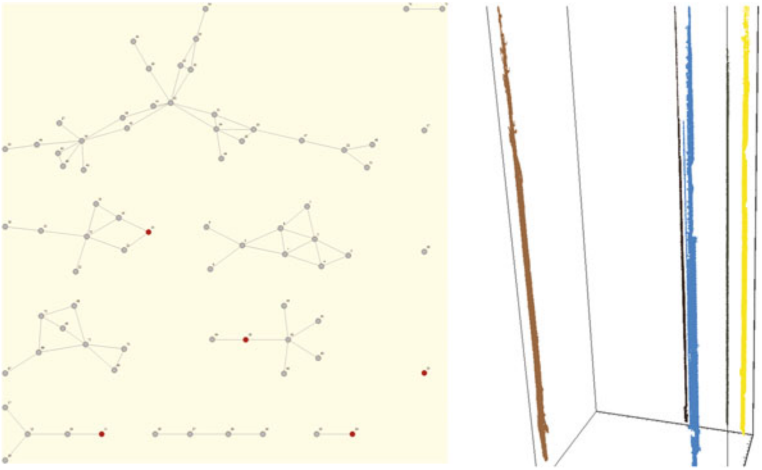


Fig. 13.6 BCGs of the whole dataset and points of five selected branches (denoted by *red* nodes in the graph)

13.3.4.1 Object and Connectivity Recognition

Object recognition is aimed at identifying and characterizing the elementary components of the complex engineering object. It processes branches of the BCG one by one. Points belonging to the voxels of a branch are taken as evidence for the existence of some specific object type. Hence, in the next steps measurement points are again directly processed. For instance, Fig. 13.6 presents the BCGs of the whole elevator dataset together with points of five selected linear branches.

First, the characteristic **axis** of a branch is found together with its start and end points. Next, the **type of the elementary object** is determined by means of bounding planes fitted to the point set. Plane fitting is executed iteratively until new planes are found with support of a predefined, minimal number of points. Planes are then projected along the axis, resulting approximate linear contours of the cross section of the supposed object. In principle this should be sufficient for identifying the shape of an extruded object, but due to noise and occlusion, in real datasets these lines are typically multiple and inaccurate. However, essential information on the type of the object can be gained by investigating the intersection of these lines and identifying its shape pattern. Hence, quasi-parallel lines running close together are substituted by a single representative, and the resulting pattern of crossing lines is processed further. After distinguishing cylindrical objects, structural elements such as pillars and beams of various profiles are sought. Their main feature is that they are linearly extruded 3D versions of some 2D linear contour. For identifying the particular object types, a **shape grammar** was developed that labels intersections of contour lines as end or middle points, and suggests a classification.

Having objects with recognized types, our earlier CAD model matching procedure is applied to determine the values of basic parameters, like center line and radius of pipes, or sizes of cuboids. Here, an iterative search maximizes the degree of match of the target object with the relevant set of points. A new method was developed for fitting various types of beams to standard elements of a catalogue. Finally, object recognition is completed with determining the local reference frames of the elementary objects. Figure 13.7 presents three recognized L-beams in the elevator data set, along with two unidentified branches.

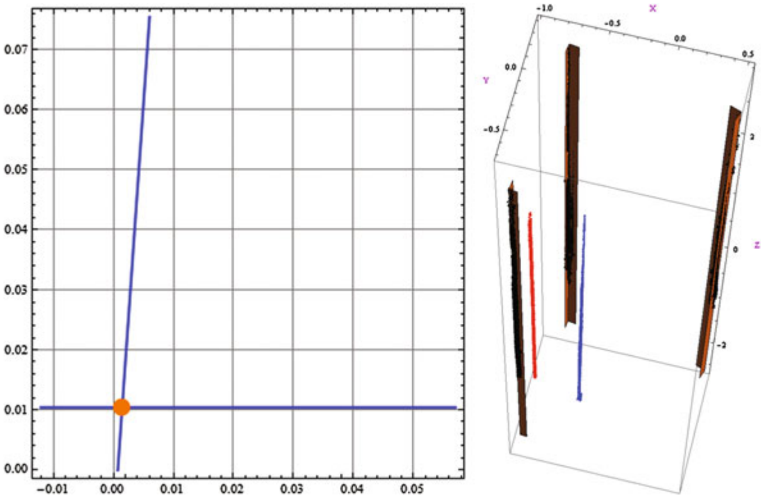


Fig. 13.7 Recognized L-beams by shape grammar. Objects behind the *red* and *blue* branches could not be recognized

In the last step, **connectivity relations** between the recognized elementary components are taken from the BCG. If the object recognition process was unsuccessful, the respective branch is labelled as unidentified. Networks of connected pipes or beams can be obtained by inducing a subgraph of given type of nodes and by determining their connected subgraphs. The final representation is rich enough for making inferences on the connectivity of different types of objects; e.g., one can deduce which beam provides support for a pipe system.

As for the performance of the overall recognition method, the 4-layered representation using points, voxels, voxel connectivity and branch connectivity graphs was powerful for recognizing both elementary components and their topological relations. The objects could be recognized even from noisy and incomplete data. In contrast to earlier approaches, our method could identify generic extruded objects, i.e., not only pipes, but also various types of beams and pillars. Routine applications in the elevator domain have shown the practical applicability, while tests run on large plant datasets have proved the scalability of the method whose performance can be improved by parallel processing.

13.4 Route Planning Technology for Replacement Task

13.4.1 Impact of Route Planning on Replacement Task

This section focuses on route planning technology for replacement task in renewal/retrofit business of plant. Insufficient route planning prevents competitive price setting that induces to lose business chance. It is important to plan accurate and efficient carry-out/in paths in short time.

The route is required to have no collision with plant structures and equipment. For the collision check, conventional 3D-CAD systems have been utilized. Engineers have to develop paths manually and the collision check on the CAD system consumes a long time. It is not a short time because it takes 5 h for one path for carrying in a boiler plant building for instance.

The manually planed carry-out/in routes are not accurate due to overlook of collision occurrences. It often turns out that the paths are not feasible on site. Besides, the manually plant route tends to be inefficient because of insufficiency of evaluation of all conditions of the task.

Above issues induce the rework of the carry-out/in tasks. The rework consumes resources of workers and equipment therefore it is one of the main reasons of incensement of the cost of replacement task.

13.4.2 Subjects of Route Planning Technology

13.4.2.1 Objective of Route Finding Algorithm

In general, route finding algorithms aim to optimize length of the route [14]. But the shortest route is not sufficient for the route in a plant building. There are two important aspects to find route for carrying a large component in/out form building for replacement task.

The one is size of corner space for direction change of the carried component. In carry-in/out work, components are suspended by overhead cranes since most components weigh more than a ton. Workers manipulate the crane to change direction of the suspended component in each corner. Each corner needs to be large enough to rotate the component without collision; the sizes of components are several meters. Therefore, largeness of each corner space on the route is an important indicator to evaluate the route. The largeness of corner space is called as ‘space margin’ from now on.

The other one is the ‘number of corners’ on the route. At each corner, workers rotate the carrying component and the posture of the carrying component is changed. Fewer corners are preferable from the view point of easiness and cost reduction of carne manipulation.

13.4.2.2 Crane Suspension Posture

Even the optimal carry-in/out route is found with respect to space margin and number of corners, feasibility of crane operation is not ensured. Collision may occur on the route. Detailed trajectory of position and posture on the route should be simulated in advance of actual crane operations. Especially, at corners, crane operations tend to be reworked because of unexpected collision occurrences.

13.4.2.3 Scale of Route Finding Problem

Most fundamental and stable algorithm for route finding is Dijkstra method [25, 8] which is applicable for many problems by designing the objective function. The fastest route finding algorithm ALT [15] is enhanced based on Dijkstra method. Lozano and Wesley introduced an idea of configuration space to route finding algorithm that enables to find posture and position sequence of the mechanism with degree of freedom.

Table 13.1 indicates computation order of individual problems. Route finding problem is represented on a graph network. Number of vertices in the graph network indicates the scale of the problem.

Table 13.1 Scale of route finding problem

No.	Problem	Vertices n	Computation order ratio O (n^2)
1	San Francisco Bay Area map [8]	321,270	1
2	Plant building (3 Dim. position)	973,210	9.2
3	Full USA map [8]	23,947,347	5,600
4	Plant building (5 Dim. posture and position)	249,141,760	600,000

No. 1 and No. 3 is ALT algorithm on road maps. Each Vertex represents a junction of roads. No. 2 is a problem which finds optimum route as position sequence in a plant building. The building 3D model is divided into small cuboids which size is less than 0.125 m^3. Each cuboid is represented as a vertex in the graph network. No. 4 expands the problem of No. 2 on order to find posture sequence on the route. Configuration space of the crane is 5 dimensions adding 2 DOF rotating along 2 axes. The configuration space expands the number of vertices as many times as the number of samples of rotated angles along two axes are. In case of No. 4, 16 × 16 samples for two axes rotation is timed to each cuboid that ends up 249,141,760 vertices in No. 4.

No. 4 problem is 600,000 times larger than the problem of No. 1. If computation time for one route finding of No. 1 is 1 min/route, then the computation time of No. 4 is assumed as about 1.1 year/route which is not feasible on business sense.

13.4.3 Break Down of Route Planning Problem

The approach of configuration space heavily consumes computation time of route finding in the plant building. To overcome enormous size of the problem, it is a practical solution that breaking down a big problem into small size problems.

The same chapters propose the breaking down approach to carry-in/out route planning [29–31]. The problem is broken down into 3 steps as below;

Step1. In 3 dimensional space of point, the optimal route is found maximizing the space margin and minimizing number of corners on the route. These indicators of the objective function of the optimal search aim to ease and increase efficiency of crane manipulation with avoiding collision.

Step2. In each corner of the found route, posture and position trajectory is simulated based on the crane suspension dynamics.

Step3. Collision between the carrying component and the building is evaluated along the route trajectory. If collision occurs, posture to avoid collision is searched in each point.

In Steps 1 and 2, the problem size becomes small enough to be computed in feasible time. But for Step 3, its problem size is still large to handle with multi core CPU. Many core of Graphic Processing Unit (GPU) has been highlighted in concurrent processing area. Therefore GPU is utilized for collision check in Step 3.

13.4.3.1 Route Finding Algorithm for Carry In/Out Plant Building

Figure 13.8 represents basic idea of route finding algorithm for carrying component in/out of plant building.

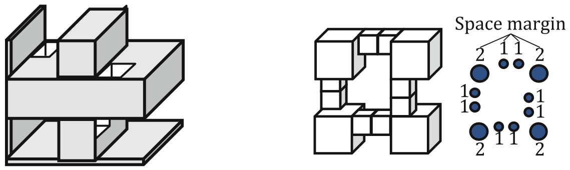
(a) Geometric 3D model of the plant building.
(b) The 3D model is divided into many cuboids called voxels. This process divides a voxel into eight voxels (i.e., octree) if the voxel includes building structures and size of the voxels are more than predefined minimal size.
(c) All voxels are divided if the size is more than predefined minimal size. In every voxel, the objective function is evaluated and its value is linked to the voxel. Here, the value of space margin is linked. Network graph is structured by voxels. In the network graph, a node indicates a voxel and an arc indicates adjacency between two voxels.
(d) The first path is found by Dijkstra's algorithm.

There are two challenges in network-graph creation for route finding.

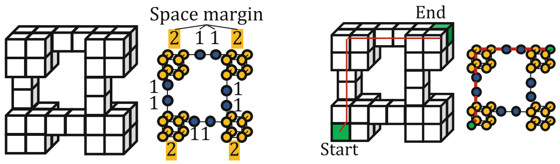
One is uniform voxel partitioning [7]. In general, octree method shown in Fig. 13.8b is used. Non-uniform voxels increase the number of corners on the route. Figure 13.9a quad-tree division tends to have more corners than uniform division in (b).

The other is vertices of network-graph has to be expanded to count number of corners [2]. Figure 13.10 shows the expansion of the network-graph. Notice at D vertex in Fig. 13.10a, there is no way to distinguish route C–D–F and route B–D–F

Fig. 13.8 Route finding algorithm for replacement task

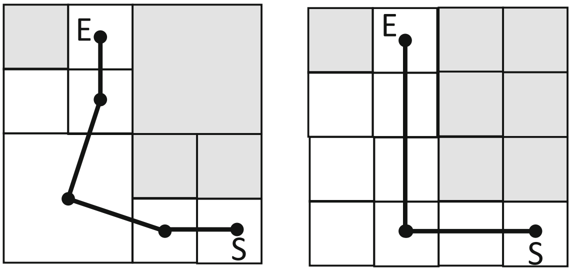


(a) 3D model of plant building **(b) Free space octants**



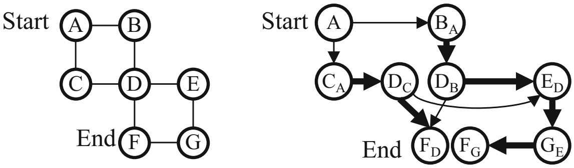
(c) Network - graph **(d) Optimal route finding**

Fig. 13.9 Number of corners depends on space division method



(a) Quadtree cells in 2D (b) Uniform cells

Fig. 13.10 Expansion of the network-graph



(a) Basic graph (b) Extended graph

since D only memorizes minimum value of number of corners. If D is duplicated as the direction to come like Fig. 13.10b, then D_C has 2 corners and D_B has 1 corner so that B–D_B–F has fewer corners than C–D_C–F.

13.4.3.2 Simulation of Crane Suspension Posture

A crane suspension is modeled as the kinematics shown in Fig. 13.11. Two chain blocks are modeled as prismatic pairs which simultaneously expand length of two wires then the posture of suspended component is changed.

Suspended angles and wire tensions indicated in Fig. 13.12 that is important for safe manipulation. Operation sequence of crane is planed then simulated checking collision and safety. Position and posture trajectory is derived from integration of the acceleration solved dynamics model expressed as differential algebraic equations [7].

13.4.3.3 Concurrent Collision Check Process with GPU

Collision check is time consuming geometric computation process. It is possible to handle collision check process independently for each pose of the suspended component. Therefore, concurrent processing on GPU is an effective solution for this problem. Collision evaluation of each interpolated trajectory pose is concurrently computed on GPU [20].

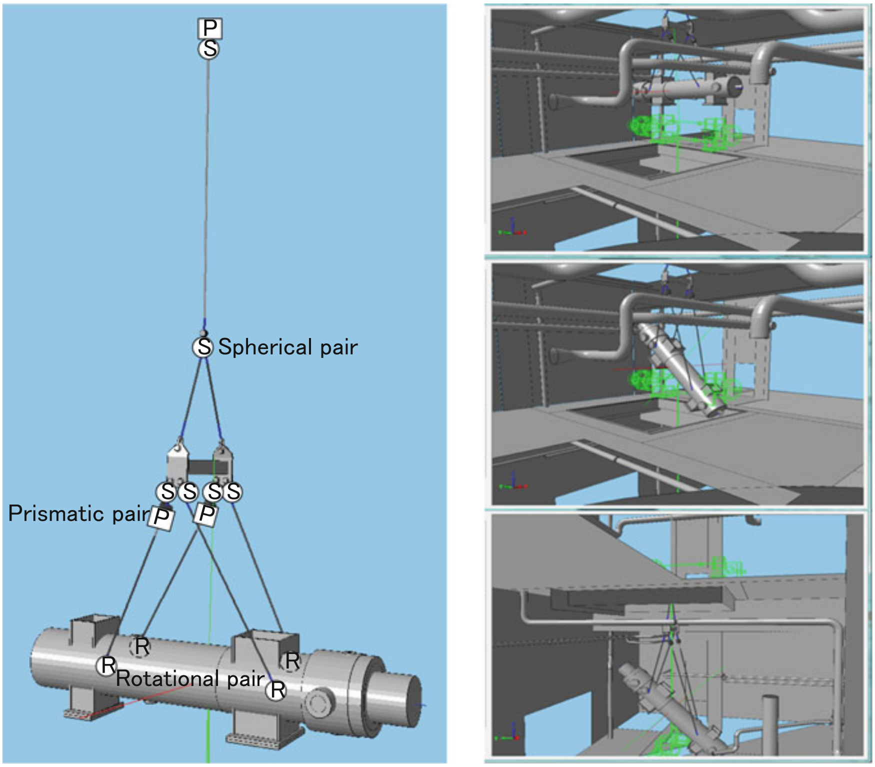
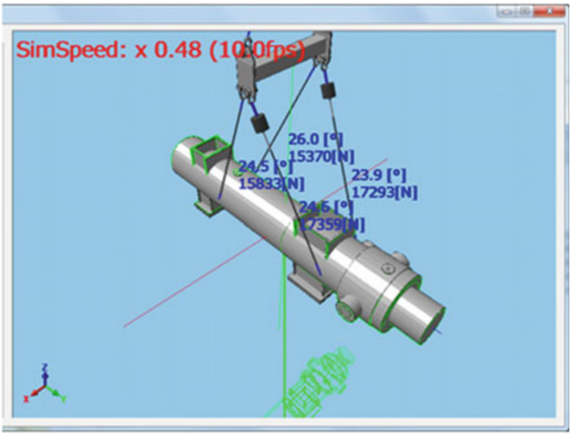


Fig. 13.11 Crane suspension dynamic model

Fig. 13.12 Suspension angles and wire tensions



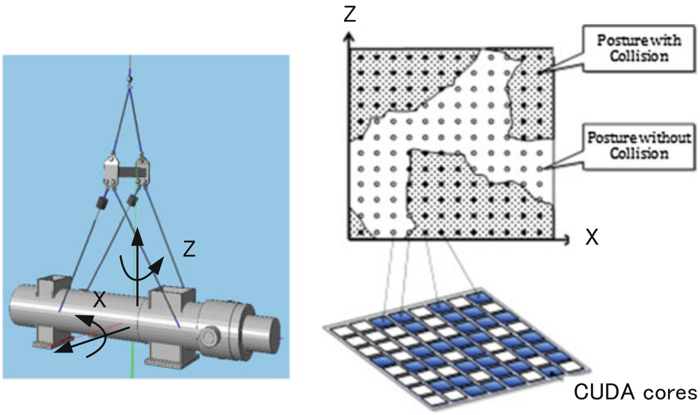


Fig. 13.13 Collision check with CUDA core

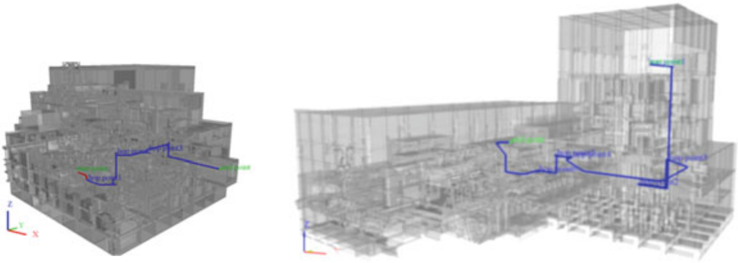

(a) Plant A (b) Plant B

Fig. 13.14 Acquired routes

In Fig. 13.13, white circles indicate postures without collisions, and the black circles indicate postures with collision. Each sampling posture rotated around x and z axis is evaluated collision occurrence by each arithmetic processing unit called CUDA core® on GPU. They are processed concurrently therefore collision check computation time decreases largely.

13.4.3.4 Computation Time of Route Planning

Sampled two plants shown in Fig. 13.14, average computation times are verified as shown in Table 13.2.

The time for a route includes route finding, position and posture trajectory generation and collision check. In both cases, it is feasible to apply to the planning of replacement tasks.

Table 13.2 Average computation time

Plant	Vertices in graph network	Average computation time (sec)
A	262,144	13.7
B	2,097,152	142.5

13.5 Conclusions

This chapter has looked at some of the retrofit and replacement technologies we developed that are driven by the latest IT. These technologies are used for reverse engineering in conformance with site surveys, and for preliminary engineering done for construction work using plant models. These technologies are now being applied to elevator systems and thermal power plants, and trial use has started for substation replacement projects.

As IT functions become more advanced, recognizing worker behaviours in addition to objects will become practical, and it will be important to manage the progress of complex retrofit/replacement projects in real-time with IT systems. These advances will enable higher utilization rates and longer equipment life, enabling highly efficient construction and maintenance of safe and reliable social infrastructure platform. We will continue to develop technologies to meet this objective.

Acknowledgements The Hungarian authors thank for the support of the Hungarian Scientific Research Fund (OTKA), Grant No. 113038.

References

1. Adolphy S, Grosser H, Kirsch L, Stark R (2015) Method for automated structuring of product data and its applications. Proc CIRP 38:153–158
2. Alice O, Atsuko E, Norisuke F, Yoichi N, Juergen R, Sonja S, Michael E, Joerg K (2015) Automatic route-finding with non-monotonic-trend factor for large-scale plant maintenance task. In: Proceedings of 48th CIRP conference on manufacturing systems—CIRP CMS 2015
3. Arayici Y, Hamilton A, Gamito P (2006) Modelling 3D scanned data to visualise and analyse the built environment for regeneration. Surv Built Environ 17(2):7–28
4. Bosché F (2010) Automated recognition of 3D CAD model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction. Adv Eng Inform 24(1):107–118
5. Cai P, Cai Y, Chandrasekaran I, Zheng J (2016) Parallel genetic algorithm based automatic path planning for crane lifting in complex environments. Autom Constr 62:133–147
6. Chang H, Li TY (1995) Assembly maintainability study with motion planning. In: Proceedings of the 1995 IEEE international conference robotics and automation, vol 1, pp 1012–1019

7. Enomoto A, Fujii N, Nonaka Y, Kolibabka J, Rasch J, Schulte S, Engelhardt M, Kaneko J, Ichijo T, Shibuta K (2014) Multiple path finding system for replacement tasks. In: 9th CIRP ICME
8. 9th DIMACS implementation challenge—shortest paths. http://www.dis.uniroma1.it/challenge9
9. Erdős G, Nakano T, Váncza J (2014) Adapting CAD models of complex engineering objects to measured point cloud data. CIRP Ann Manuf Technol 63(1):157–160
10. Erdős G, Nakano T, Horváth G, Nonaka Y, Váncza J (2015) Recognition of complex engineering objects from large-scale point clouds. CIRP Ann Manuf Technol 64(1):165–168
11. Farnsworth M, Tomiyama T (2014) Capturing, classification and concept generation for automated maintenance tasks. CIRP Ann Manuf Technol 63(1):149–152
12. Fathi H, Dai F, Lourakis M (2015) Automated as-built 3D reconstruction of civil infrastructure using computer vision: achievements, opportunities, and challenges. Adv Eng Inform 29(2):149–161
13. Geng J, Tian X, Bai M, Jia X, Liu X (2014) A design method for three-dimensional maintenance, repair and overhaul job card of complex products. Comput Ind 65(1):200–209
14. Ghandi S, Masehian E (2015) Review and taxonomies of assembly and disassembly path planning problems and approaches. Comput Aided Des 67:58–86
15. Goldberg AV (2005–2010) Efficient point-to-point shortest path algorithms; technical report MSR-TR-2005-132
16. Hsu D, Kavraki LE, Latombe JC, Motwani R, Sorkin S (1998) On finding narrow passages with probabilistic roadmap planners. In Robotics: the algorithmic perspective: 1998 workshop on the algorithmic foundations of robotics, pp 141–154
17. Lei Z, Han S, Bouferguène A, Taghaddos H, Hermann U. Al-Hussein M (2014) Algorithm for mobile crane walking path planning in congested industrial plants. J Constr Eng Manag 141(2):05014016-1-10
18. Németh I, Püspöki J, Cs Haraszkó, Gy Mátyási, Nagy T, Freeman Ch, Scott RW, Baldwin JS (2013) 3D design support for rapid virtual prototyping of manufacturing systems. Proc CIRP 7:431–436
19. Nonaka Y, Enomoto A, Fujii N, Kolibabka J, Rasch J, Schulte S, Engelhardt M, Kaneko J, Ichijo T (2014) Fast path finding system with GPGPU computing for replacement tasks in plant maintenance. Proc CIRP 22:23–32
20. Nonaka Y, Enomoto A, Fujii N, Kolibabka J, Rasch J, Schulte S, Engelhardt M, Kaneko J, Ichijo T, Shibuta K (2014) Fast path finding system with GPGPU computing for replacement tasks in plant maintenance. Proc CIRP TESCONF 22:23–32
21. Ochmann S, Vock R, Wessel R, Klein R (2016) Automatic reconstruction of parametric building models from indoor point clouds. Comput Graph 54:94–103
22. Oesau S, Lafarge F, Alliez P (2014) Indoor scene reconstruction using feature sensitive primitive extraction and graph-cut. ISPRS J Photogramm Remote Sens 90:68–82
23. Pătrăucean V, Armeni I, Nahangi M, Yeung J, Brilakis I, Haas C (2015) State of research in automatic as-built modelling. Adv Eng Inform 29(2):162–171
24. Schön B, Mosa ASM, Laefer DF, Bertolotto M (2013) Octree-based indexing for 3D pointclouds within an Oracle spatial DBMS. Comput Geosci 51:430–438
25. Schrijver A (2002) Combinatorial optimization, vol A. Springer, pp 96–106
26. Son H, Bosché F, Kim C (2015) As-built data acquisition and its use in production monitoring and automated layout of civil infrastructure: a survey. Adv Eng Inform 29(2):172–183
27. Stark R, Grosser H, Müller P (2013) Product analysis automation for digital MRO based on intelligent 3D data acquisition. CIRP Ann Manuf Technol 62(1):123–126
28. Tang P, Huber D, Akinci B, Lipman R, Lytle A (2010) Automatic reconstruction of as-built building information models from laser-scanned point clouds: a review of related techniques. Autom Constr 19(7):829–843

29. Wang C, Cho YK (2015) Smart scanning and near real-time 3D surface modeling of dynamic construction equipment from a point cloud. Autom Constr 49:239–249
30. Wang J, Gu D, Yu Z, Tan C, Zhou L (2012) A framework for 3D model reconstruction in reverse engineering. Comput Ind Eng 63:1189–1200
31. Xiong X, Adan A, Akinci B, Huber D (2013) Automatic creation of semantically rich 3D building models from laser scanner data. Autom Constr 31:325–337

Part V
System Degradation and Design

Chapter 14
Infrastructure/Train Borne Measurements in Support of UK Railway System Performance—Gaining Insight Through Systematic Analysis and Modelling

Amir Toossi, Lloyd Barson, Bradley Hyland, Wilson Fung and Nigel Best

Abstract The British railway is one of the most complex mixed traffic railway systems in the world. Increasing passenger/freight traffic is placing significant demand on the railway system calling for improvements to the existing network capacity and service reliability. Through systematic data analysis and modelling, the industry has gained a better understanding and insight into the system behaviour. This has highlighted the necessity for higher precision measurements of the railway system to enable an effective optimisation of capacity utilisation and reliability of service performance. This chapter describes two main root causes affecting the service performance and the measurement systems used to quantify their impact. The current performance modelling techniques based on infrastructure-borne and train-borne measurement systems and their pros and cons will be discussed. Finally, opportunities will be presented to improve the performance monitoring and measurement systems by using the various data layers available to the industry in a better managed way.

14.1 Introduction

14.1.1 Background

In 2002 Network Rail took over as owner and operator of Britain's railway infrastructure with a mandate from the British government to improve safety, reliability and efficiency on the network. Network Rail operates the railway under a network licence issued by the Office of Rail and Road (ORR) which places obligations on Network Rail to operate and maintain a railway network that delivers acceptable levels of performance to all stakeholders.

A. Toossi (✉) · L. Barson · B. Hyland · W. Fung · N. Best
Network Rail, Systems Analysis Team, Milton Keynes, UK
e-mail: amir.toossi@wspgroup.com

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_14

Since 2002, Network Rail has almost halved the cost of running the railway, driven up its asset reliability to record levels, significantly improved train punctuality and made Britain's railway the safest in Europe [1]. Over the coming years, the most urgent challenge facing the railway is the need to increase capacity. Passenger numbers have doubled over the past 20 years—more than 4.5 million people use the rail network every day. This is twice as many passengers as in the 1920s—when the network was more than twice its current size. Almost 1.7 billion passenger journeys are now made on the railway every year—a third more journeys than five years ago and this level of growth is set to continue. In 2007, the UK Department for Transport (DfT) forecasted rail passenger journeys to double in 30 years [2].

14.1.2 The Challenge

With increasing demand on the railway from both passenger and freight traffic, congestion is a huge daily challenge, particularly at peak times of day. Whilst a number of large infrastructure projects are underway to increase the capacity of the railway, there is a need to effectively plan the timetable to use the current railway capacity to deliver reliable service performance. As a consequence, significant emphasis is placed on understanding the capacity of the railway and how this capacity is utilised.

This paper focuses on the outcomes of Systems Modelling projects conducted by Network Rail's Systems Analysis Team based on utilising industry data sources to better understand the prevailing factors affecting system's performance. The paper details opportunities to analyse and model the railway system with higher levels of fidelity to ultimately support the efficient and robust allocation of railway capacity.

14.2 Regulatory Measures of Service Performance

There are two main industry standard measurements of performance—Public Performance Measure (PPM) and Delay minutes. PPM is a measure of the punctuality and reliability of passenger trains in Britain. It is the percentage of scheduled trains which successfully run their entire planned route, calling at all timetabled stations, and arrive at their terminating station 'on time', where 'on time' means within 5 min of the scheduled destination arrival time for London and South East and regional operators (i.e. commuter services), or within 10 min for long-distance operators. The target for PPM varies each year and for each train operator, and is agreed with the ORR. PPM measures the performance of individual trains advertised as passenger services against their planned timetable as agreed between the operator and Network Rail at 22:00 the night before. PPM is therefore the percentage of trains 'on time' compared to the total number of trains planned.

Delay minutes are a performance measure for punctuality of passenger and freight trains. A delay is defined as a loss of time against schedule between two consecutive locations on the train service journey. The delay minutes are attributed to the cause of the initial delay where a service incurs delays of 3 min or more between the specified locations:

- Network Rail caused delays: as well as infrastructure faults this figure includes external factors such as weather, trespass, vandalism, cable theft and fatalities.
- Train operator caused to self: delays to a passenger train operating company's services that are caused by that company.
- Caused by other train operators: delays to a passenger train operators services that are caused by another train company.

14.3 Factors Affecting Service Performance

Timetable Planning Rules (TPRs) form the basis for allocating the railway capacity to the services which run on the network. These rules are typically derived from modelling of the infrastructure (e.g. signalling configuration) and rolling stock performance. On an ideal railway, the timetable would comply with these Timetable Planning Rules and the infrastructure and operational functions would exhibit perfect reliability such that services would always arrive on time. In reality, the reliability of railway system functions is not perfect and the associated incidents can have a significant impact on service performance. These incidents fall into two primary categories.

14.3.1 Special Cause Incidents

Low-frequency, high-impact events which are inherently unpredictable and are typically assignable within the system to a root cause e.g. Points failure, Extreme weather. These incidents are captured in the Network Rail TRUST system (Train Running System on TOPS (Total Operations Processing System)) which records detailed information about the service-affecting failure modes. These incidents are only recorded if the delay impact on services is greater than 3 min. Some examples of special-cause variation are:

- Trespass and Vandalism
- Points failure
- Extreme weather

14.3.2 Common Cause Variations

High-frequency, low-impact events that are constantly active within the system and are typically unassigned within the system to a root cause. These are incidents that affect the system and cause delays less than 3 min duration. They take a great deal of effort and cost to capture them in the current failure recording systems in industry. The results of the Brighton Mainline reliability modelling (BML) study (T1019) [3], highlighted the importance of reducing the common-cause variations to build a more robust system with higher levels of service performance.

Some examples of common-cause variations are:

- Constraining service interaction planned into timetable i.e. technical headways and junction margins.
- A legacy set of Timetable Planning Rules, not fit for purpose.
- Variance in driver behavior.
- Minor station dwell time overrun.
- A vehicle fault occurring in many units resulting in variance in tractive effort performance across a vehicle class.

As shown in Fig. 14.1, in simple terms, the railway system operation can be represented by four main activities: Planning the timetable, signalling the railways, driving the trains and measuring the performance based on predefined performance indicators. Historically industry has been working effectively to capture the Special Cause events. However, we need better ways to integrate the data sources and benefit from the data that provides us a great insight into the Common Cause events and their underlying factors. The N^2 plot demonstrated in Fig. 14.1, highlights the importance of using the available data sources through a rigorous process improve the system performance in its dynamic operational environment.

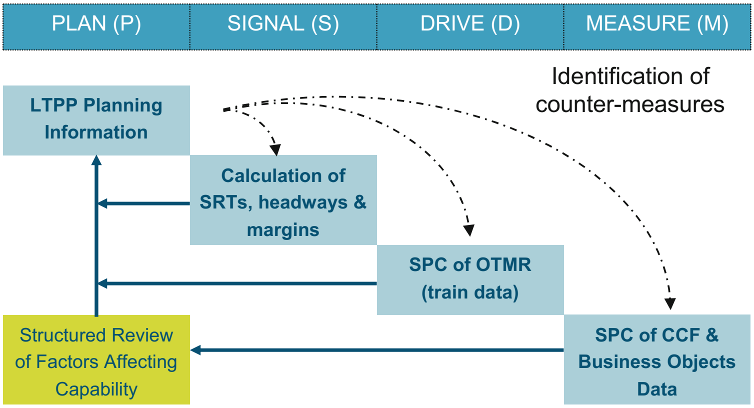


Fig. 14.1 Simple N^2 plot for the railway system design

14.4 Railway Systems Modelling and Analysis

Models are abstractions of reality and have inherent limitations, some of which are emergent. However, modelling can provide insight when these limitations are understood, models are built to standards, inputs are industry aligned and endorsed, and models are frequently validated against the best model available, the real railway.

There are a number a tested and validated modelling tools and methodologies available across the industry. However, for the purpose of this chapter, we focus on four main categories:

- Capacity Modelling
- Journey Time Modelling
- Performance Modelling
- On-Train Measurement Recorder (OTMR) data analysis.

These modelling and analysis techniques are further described in the following sections.

14.4.1 Capacity Modelling

The railway capacity defines the quantity of trains that can travel over a specified route during a particular time period. Signalling Performance Assessments (SPA) are used to calculate the capacity of the railway which is defined by the following values:

- Technical Headway: The Technical Headway is specified at a signal location and defines the minimum time interval between two trains passing that location such that the following train observes an unrestrictive aspect (i.e. green signal aspect).
- Junction Margin—Defines the minimum time interval between two trains passing a specified Timing Point Location (TIPLOC) such that the following train observes unrestrictive aspects.
- Platform Reoccupation—the Platform Reoccupation is specified at a TIPLOC and defines the minimum time interval between the departure of a train from a platform and the arrival of a train to the same platform.

The SPA uses outputs from the Network Rail VISION software tool which simulates train movements across a modelled railway layout—this model captures the track layout, signalling elements, gradients, rolling stock performance and driver behaviour. Variation in driver behaviour, such as braking rates, can have a significant impact on capacity and therefore actual measurements can provide evidence to quantify the impact.

14.4.2 Journey Time Modelling

This model calculates minimum running times between TIPLOCs for specified train configurations. Train timing data is used to validate these models.

The Operational Planning discipline within Network Rail is responsible for allocating capacity for passenger and freight services. Using the outputs from SPA and Journey Time Modelling, Timetable Planning Rules (TPRs) are created which form the basis for allocating the railway capacity to the services which run on the network.

14.4.3 Performance Modelling

The Network Rail Systems Analysis Team uses the TRAIL (Transport, Reliability, Availability and Integrated Logistics) software tool to carry out performance modelling of the railway system—this modelling is primarily used to forecast the performance of future railway states. TRAIL is a discrete event simulator which models the following elements of the railway system:

- Infrastructure: tracks, switches and crossings, and signals
- Timetable: services and their associated timings
- System Reliability: railway systems functional failure modes and their impact on service performance
- Operations: contingency plans e.g. cancellation thresholds.

The model uses random sampling to generate system failures and the associated impacts. The model is built based on the observed data from industry business systems, and input from route technical and operations experts. It seeks to reflect any incidents that could impact upon services. A method to apply common cause variance in the system performance model was developed through the Brighton Mainline Reliability Modelling (BML) Study [3] and this is now used in production models. The method is designed to reflect both the frequency of common cause events as well as the calculated impacts, as reduced running times. A summary of the outcomes of the Brighton Mainline Modelling Study which provided novel insights into the optimal means to improve system's performance is provided in the following sections.

14.4.4 BML Baseline Performance Model Results

The industry believes that Brighton Mainline is being operated beyond the capacity limits of the infrastructure hence a study was commissioned by RSSB to investigate the factors affecting its performance. The study looks at the performance of the

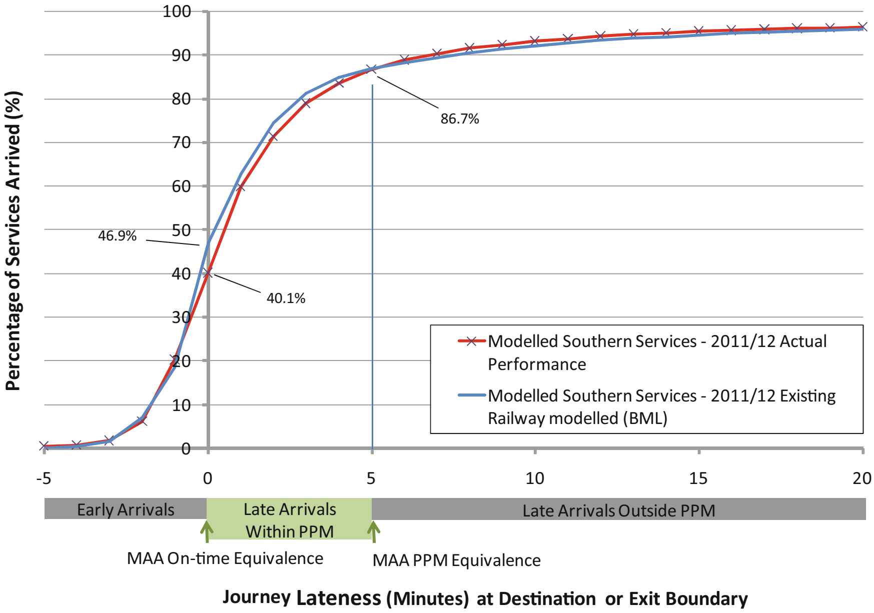


Fig. 14.2 Modelled and actual journey lateness 2011/12

BML during the 2011/12. The results of this modelling study are presented in this section. Figure 14.2 is a journey lateness curve showing the cumulative percentage of services arrived against journey lateness. It compares the output of the baseline 2011/12 BML model to the observed BML data from 2011/12. The alignment between the series is good overall and at the point of MAA PPM equivalence it is 86.7%. The alignment at the point of on-time indicates that there is a shortfall in the evidence generated by the application of common cause impacts. This is a small shortfall relatively and felt to be acceptable for the purposes of the study.

The overall industry PPM target for England and Wales is currently set at 92.5% MAA PPM to be achieved by March 2019 [4].

After competition of the modelling work, a range of sensitivity analysis is preformed to better understand the factors affecting the service performance. These results are presented in the following section.

14.4.4.1 Sensitivity Testing Results

The key control factors were sensitivity tested by independently adjusting the input parameters in the model, simulating, and comparing the change in output in terms of service performance Fig. 14.3 shows an example of sensitivity testing output for the Infrastructure category, as output PPM Equivalence values against the input reliability, in Mean Time Between Service Affecting Failure (MTBSAF). Optimal

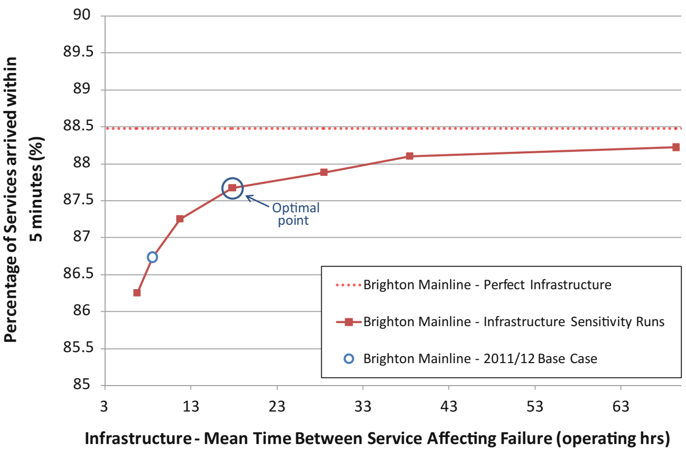


Fig. 14.3 Modelled infrastructure sensitivity analysis

reliability inputs are identified at points where increasing the reliability further produces a limited improvement in system performance and requires significant effort and cost to achieve. This approach assumes that the costs of achieving reliability are linear with MTBSAF [5].

The optimal reliability points were identified for the following special cause control factors: infrastructure; operations; weather, trespass & vandalism; operations; rolling stock. Figure 14.4 shows the result of implementing all special cause optimal reliabilities together. These combined changes yield a 5.3% lift in MAA PPM, which is significant as this equates to 40% of the unreliability in the system. The lift of 5.1% in on-time performance is small when considered the same way. It should be noted that this would be challenging and expensive to achieve these optimal reliabilities in practice with the targets being ambitious when compared to reliability levels currently being achieved.

The common cause inputs (critical system variable) were assessed by removing their impact completely from the base case. This was to establish their performance improvement potential on the BML. Figure 14.5 shows the impact of the categories independently.

Figure 14.6 shows the simulated impact of collectively removing these categories. In practice this reflects a BML railway being planned and operated to be compliant with the industry agreed Timetable Planning Rules (TPR), increased station management, with trains driven consistently to the technical capability of the rolling stock. It shows a significant lift of system performance at both the MAA on-time and MAA PPM equivalence.

This study demonstrated that Common Cause events have a significant influence on service performance (achieving PPM) and therefore it is important to better understand these events occurring on the actual railway system. The following sections describe efforts to improve the industry's understanding of Common Cause events by drawing upon data from both infrastructure and train borne measurement

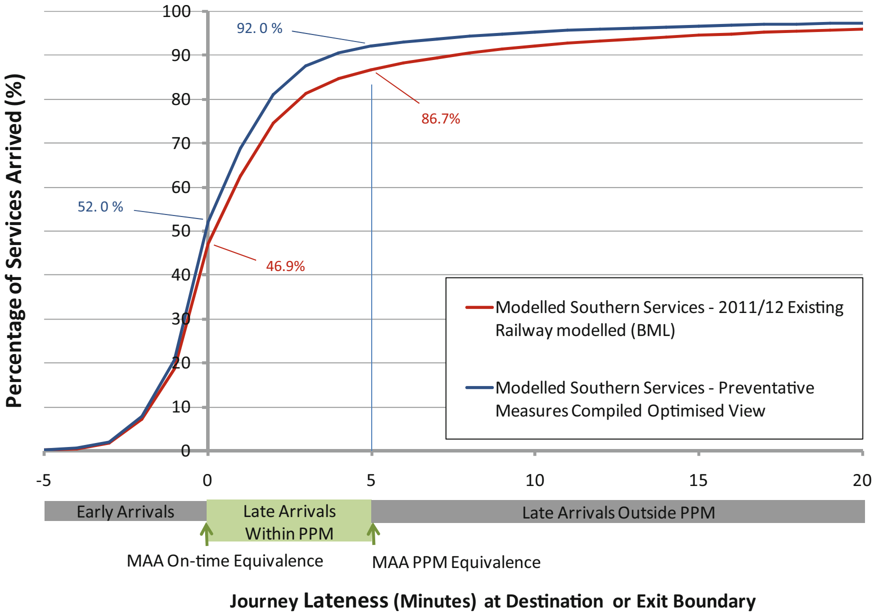


Fig. 14.4 Special cause sensitivity analysis

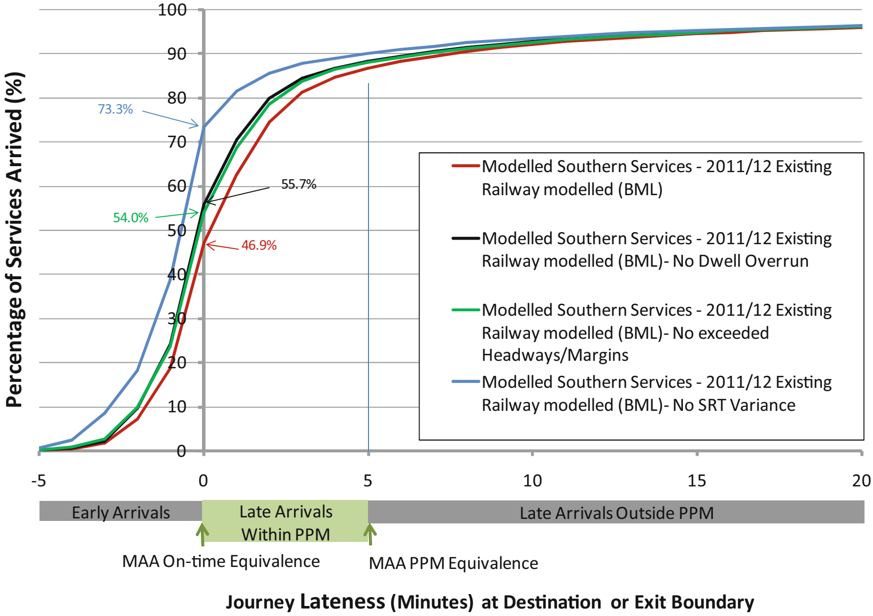


Fig. 14.5 Common cause sensitivity analysis (independent cases)

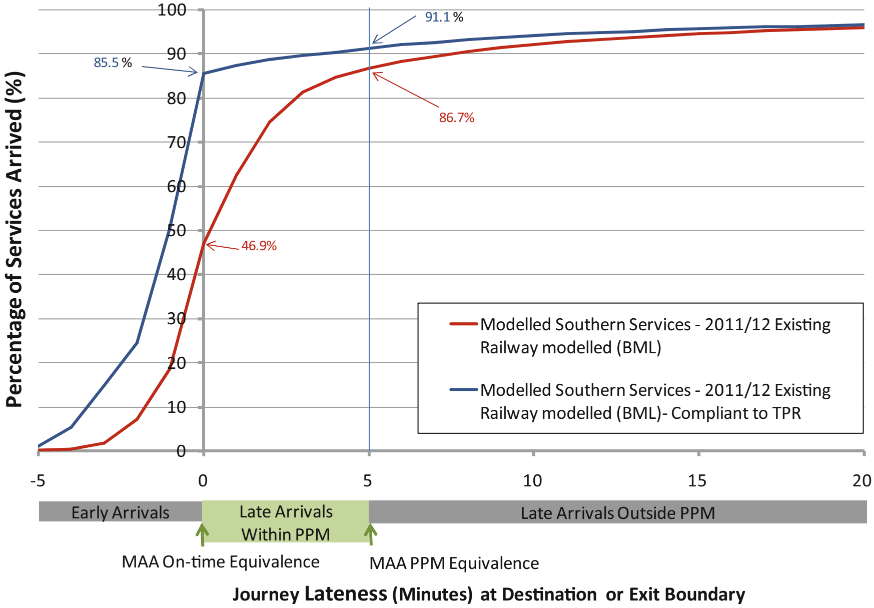


Fig. 14.6 Common cause sensitivity analysis (timetable planning rule compliance case)

systems. Also, On-Train Measurement Recorder (OTMR) and GPS data analysis will be introduced which paves the way towards a better understanding of the railway system.

14.5 Infrastructure-Borne Performance Measurement Systems

14.5.1 TRUST (Train Running System on TOPS (Total Operations Processing System)

The TRUST system is used to record the time at which a train service arrives at, departs from or passes specified Timing Point Locations (TIPLOCs); this is the primary system used for performance monitoring of services operating on the railway network. Each of the timing points specified in TRUST may fall into one of the following categories:

- Contractual Monitoring Point: the lateness measured at these locations and is used to calculate any financial penalties arising from delays caused by Network Rail or the TOCs/FOCs.
- Delay Recording Point: non-contractual, used to record delays
- Timing Point: used for monitoring.

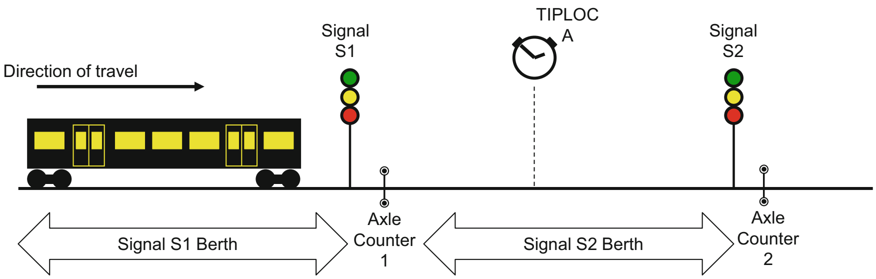


Fig. 14.7 Infrastructure elements involved in service timing

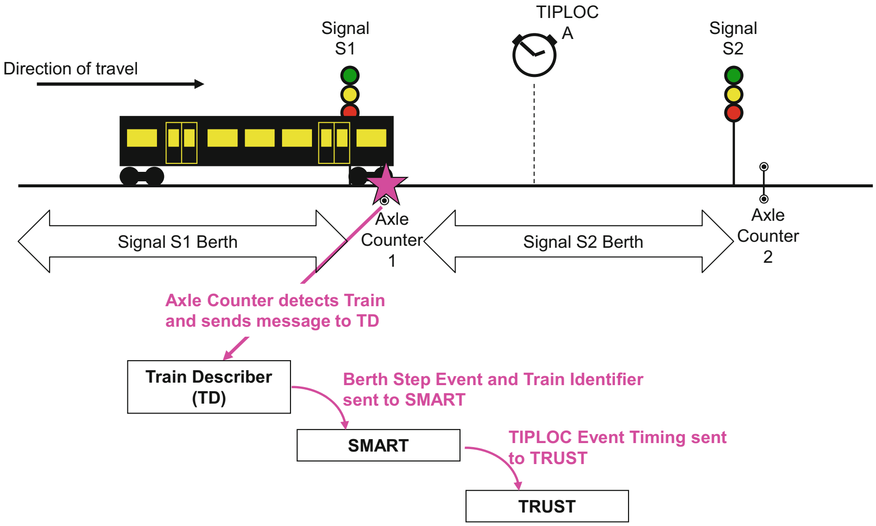


Fig. 14.8 Service timing event process

The TRUST timings are obtained using train detection systems on the Network Rail infrastructure as illustrated in Figs. 14.7 and 14.8. In Fig. 14.7, axle counters are used to detect trains entering or leaving a particular section of track—the number of axles is counted as the train passes the sensor and this information is used to determine whether the track section is occupied or clear. A Berth is normally associated with the track section which will hold the train when the signal is displaying a red aspect (an example layout showing berth moves can be found at [6]. The timing point location (TIPLOC “A”) in Fig. 14.7 is positioned between signals S1 and S2 within the signal S2 Berth.

In Fig. 14.8, the time of the train passing TIPLOC A would be obtained as follows:

- As the train reaches Axle Counter 1 and moves ("steps") from the Signal S1 berth into the Signal S2 berth, the detection event is sent to the Train Describer. The Train Describer associates the berth "step" with the train identifier and this information is then passed to the SMART system.
- The SMART (Signal Monitoring and Reporting to TRUST) calculates the timing at TIPLOC A based on the time of the train "step" event from the S1 berth to the S2 berth. In the example shown, the TIPLOC is not coincident with the axle counter location and therefore a time offset (known as a berth offset) is added to the time of the berth "step" event. This offset, which is fixed for a given berth "step", is calculated by obtaining timings of services passing the actual location (via independent timing systems) and comparing these measurements against the SMART timings for the associated berth "step".
- The SMART system passes the TIPLOC pass event time to TRUST which records only the hour and minute portion of the timing—i.e. the seconds are ignored.

Whilst TRUST timing data is readily available from the Network Rail Business systems, the data itself has some drawbacks with regard to its use in railway system modelling:

- The difference between TRUST recorded and SMART calculated timings can be up to 59 s. The lack of resolution in TRUST data can result in unrepresentative TIPLOC to TIPLOC timings, particularly where these locations are close to one another—for example, some TIPLOC to TIPLOC moves can take less than 60 s to complete.
- The TRUST timings are only taken at certain locations which may not include all the TIPLOCs used to plan the train services (in the Working Timetable).

14.5.1.1 Control Centre of the Future (CCF)

The Control Centre of the Future (CCF) system receives data from the Train Describer and displays real time train service positions on a track level map as shown in Fig. 14.9. The CCF system also logs the berth "step" events which can be replayed at a later time (e.g. when investigating delays) or exported to text files for further analysis. The exported CCF data includes the following information about each "step" event:

- From Berth (which the train has vacated)
- To Berth (which the train has entered)
- Step Event Time (hh:mm:ss)
- Train Service Identifier (Headcode)

Using the CCF log data, it is possible to calculate TIPLOC times (including seconds) from the relevant berth "step" events after applying the appropriate berth offset values.

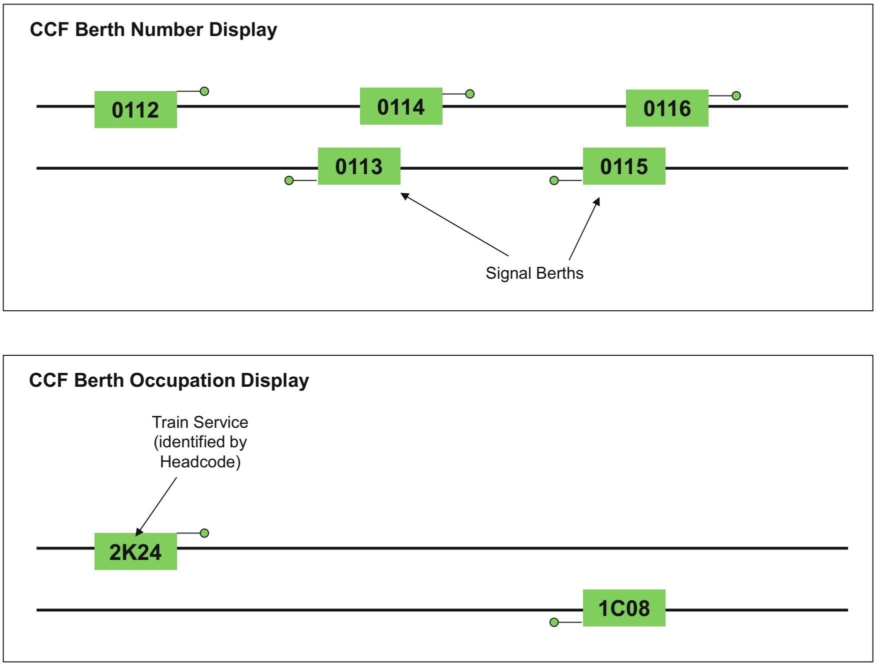


Fig. 14.9 CCF display formats

Timings derived using CCF data:

- Provide higher resolution TIPLOC timings than available from TRUST—i.e. includes seconds.
- For a given train service, timings are available at every berth traversed rather than TIPLOCs which tend to be separated by larger distances.

14.5.1.2 Berth Offsets

TIPLOC timings obtained from the TRUST system and derived using CCF data are both dependent on berth offsets where the TIPLOC is not coincident with a berth boundary. The berth offset value is fixed for a given berth step resulting in TIPLOC timings being subject to the following potential errors:

- A section of track is typically used by more than one rolling stock type and therefore the berth offset value is a compromise due to the differing performance of these trains. In Fig. 14.10, Trains A and B depart TIPLOC “A” but train B exhibits better acceleration performance than train A. Upon stepping into the berth associated with the TIPLOC “B”, train B reaches the TIPLOC in less time than train A and would therefore require a lower berth offset value.

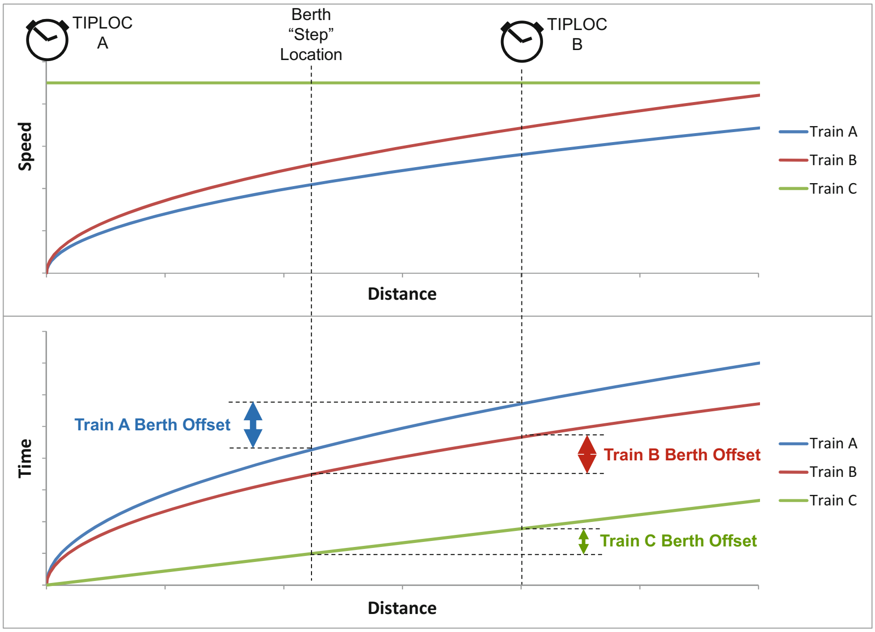


Fig. 14.10 Train performance and stopping pattern impact on berth offsets

- The stopping pattern of services passing the TIPLOC may vary which can impact the speed profile within the TIPLOC berth section. In Fig. 14.5, train C does not stop at TIPLOC "A" and would exhibit a lower berth offset time (when timing at TIPLOC "B") than trains A and B.

The speed profile of services within the TIPLOC berth will be subject to variation based on a number of factors including:

- Driving Style
- Railhead conditions (poor adhesion)
- Restrictive (non-green) signal aspects being displayed
- Temporary Speed Restrictions in force.

14.6 Train-Borne Measurement Systems

14.6.1 OTMR (On-Train Monitoring and Recording)

Trains running on the Network Rail infrastructure are required to carry On-Train Monitoring and Recording (OTMR) equipment for the purpose of:

- Providing information regarding the train system state leading up to and during any incident/accident.

- Monitoring the performance of the train vehicle system and the driver.

The relevant Railway Group Standard (GM/RT 2472) specifies a minimum set of parameters which must be recorded by the OTMR system whilst the train is in operation including speed, distance, traction demand, braking demand and the state of various protection systems (e.g. Train Protection and Warning System (TPWS), passenger alarms).

The OTMR data is typically downloaded from a train vehicle during maintenance and parameters can be viewed using dedicated data processing software—e.g. Arrowvale, QUADS. These parameters can be exported into text files (e.g. Comma Separated Variable) for analysis in packages such as Excel and MATLAB.

Compared to the infrastructure based measurements, OTMR data offers advantages including:

- Distance (used to derive position) for every time sample (instead of at fixed locations such as berth boundaries (CCF) or TIPLOCs (TRUST)). Berth offsets are not required.
- Sampling frequency typically higher than 1 Hz.
- Additional parameters defining the train state—e.g. speed, acceleration, passenger doors status.
- Additional parameters defining the driver behaviour—e.g. traction demand, brake demand.

The distance parameter in OTMR data is equivalent to the odometer reading in a typical road vehicle and therefore does not specify the actual position of the train on the rail infrastructure (unlike the berth “step” events in CCF log data). When using OTMR data for modelling, it is necessary to align the distance values to locations on the rail infrastructure as illustrated in Fig. 14.11.

The stopping pattern of the service contained in the OTMR data is extracted from TRUST along with actual timings of each depart and arrive event (as recorded by TRUST). Depart and arrive events are identified in the OTMR data based on the speed signal increasing from or decreasing to zero. The stopping TIPLOCs can then

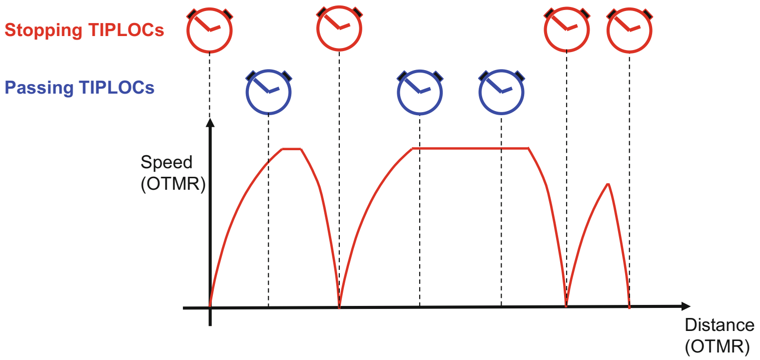


Fig. 14.11 TIPLOC locations aligned to OTMR speed versus distance data

be identified by mapping the OTMR depart/arrive events timings to the TRUST depart/arrive event timings.

Once the stopping TIPLOCs have been identified in the OTMR data and knowing the relative distances between all the TIPLOCs, the timings at “passing” TIPLOCs can be determined from the OTMR distance parameter.

In addition to extracting TIPLOC timings, the OTMR speed versus distance profile can be compared against VISION calculated (deterministic) profiles such as the station platform arrival shown in Fig. 14.12. The deceleration profile does impact the capacity of the railway system and therefore OTMR data can provide a valuable insight into the impact of variation in driving style.

14.6.2 GPS Measurements

As part of the Network Rail Timetable Rules Improvement Programme (TRIP), Global Positioning System (GPS) derived train position data was obtained for certain rolling stock types. Compared to conventional OTMR data, the GPS position information is absolute (latitude, longitude) rather than relative which significantly improves the time taken to identify TIPLOC timings. However, some drawbacks were evident whilst processing this data:

- GPS measurements were not available where the train was travelling through tunnels
- The relatively low sampling frequency (wrt OTMR) of GPS position/speed data prohibited its application to braking rate analysis.

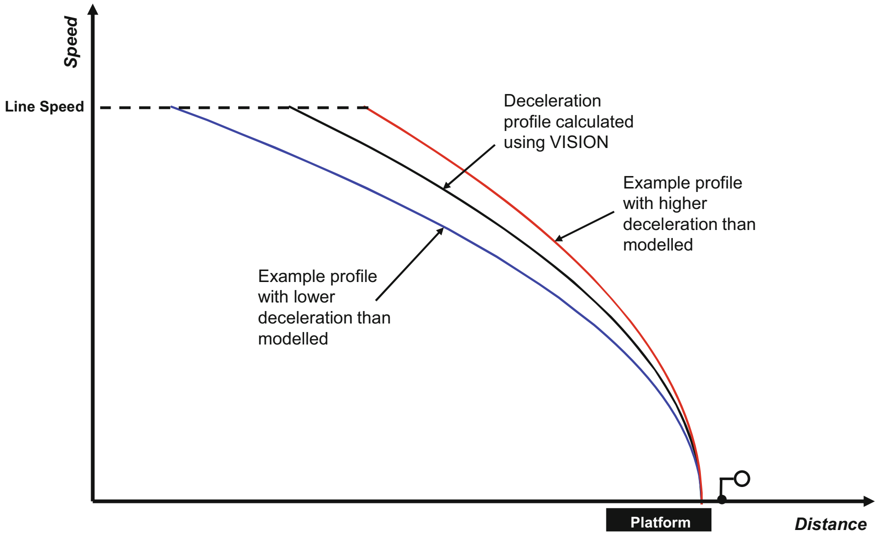


Fig. 14.12 Variation in braking profile relative to SPA (VISION) Modelled

14.7 Case Study: Northern Rail GPS Tracking System

Providing accurate train running information can be difficult because the signalling on some UK routes often doesn't have the technology to automatically report a train's location.

Northern Rail is one of the first operators that have equipped their rolling stock fleet with the GPS tracking system. The equipment has been supplied by Nomad Digital and has been partly funded by Network Rail. Customers on this particular route are now benefiting from more accurate travel information with a trial of GPS vehicle tracking capabilities on Northern trains.

GPS tracking equipment is fitted to the train and uses commercial mobile phone networks to regularly report back the position of the rolling stock to the Northern Rail's control centre. This means that the operator can provide exact and up to date information about the train's location to the stations in the trial. Standard signalling equipment is very expensive to replace therefore the trial of GPS equipment could see a lower cost alternative for tracking trains. If it's successful, this solution has the potential to be used on many routes across the UK rail network.

14.7.1 Data Collection and Format

GPS data are managed by the operator and are downloadable through their system. Data can be downloaded for each particular vehicle on a specific date. Data were collected for the end of February to March 2015 which covers rolling stock classes 150, 153, 155, 156 and 158. Data are typically received in Excel format and contain 38 fields from which we normally use the following parameters:

- Unit ID to identify the exact physical vehicle
- Date/Time, which is typically reported every 10 s
- GPS latitude and longitude
- Number of GPS satellites for the accuracy of the points
- GPS speed (miles/hr).

14.7.2 Data Preparation and Analysis

In general, we need three main parameters derived from the data: time, location and speed. In terms of the location of each GPS reading point we need to identify the Track ID/Line code (e.g. Down Fast Line, Up Slow Line) and Engineering Line Reference (ELR) that the service is travelling on which is a reference code that is used to identify a length of track. For this purpose, ArcGIS has been used to map the services' GPS reading onto the geography to be able to identify the Line code

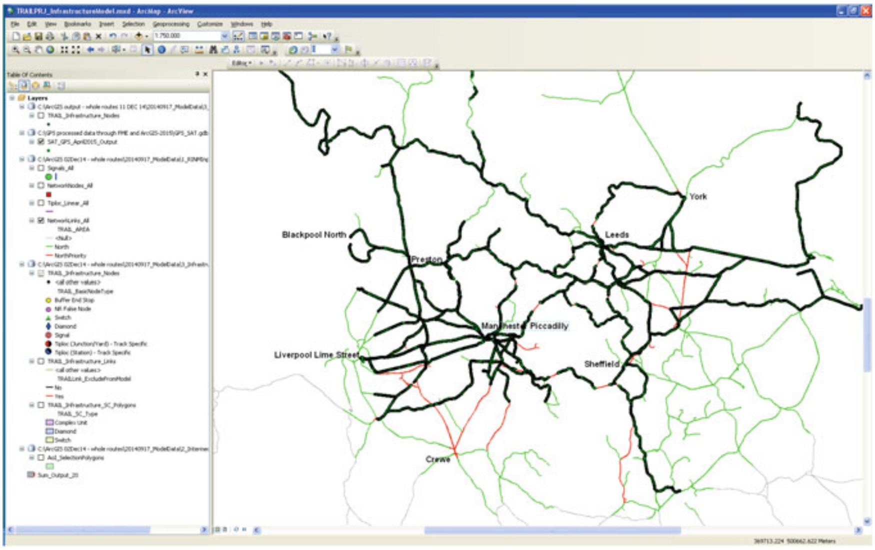


Fig. 14.13 Using ArcGIS to map the GPS readings

and ELR (see Fig. 14.13). The dark lines show us an accumulation of GPS reading for a month on the northern geography.

One of the issues with the GPS readings are the stopping services for which the GPS points start to move around that makes it difficult to identify the exact location of the train (Fig. 14.14).

After identification of the ELRs and Line codes, we also need to identify the location of the services against the Timing Point Locations i.e. stations and junctions. This step has been done through an in-house Excel VB Macro which prepares the data for the analysis.

For the purpose of this study, the GPS data were analysed to calculate the station dwell times and observed running times which more details have been described in the following sections.

14.7.3 Station Dwell Time Analysis

As previously discussed, one of the main common cause factors affecting the service performance are the minor dwell overruns affecting the stopping services at stations. In order to better investigate this impact, the station dwell times have been calculated for the stopping stations for both peak and off-peak hours. These values were then compared against the planned dwell times for each specific service.

The results for five stations in the northern geography have been demonstrated in Fig. 14.15. Box and Whisker charts have been used for this purpose to illustrate control factor variance relative to control lines. As you can see in Fig. 14.15, the

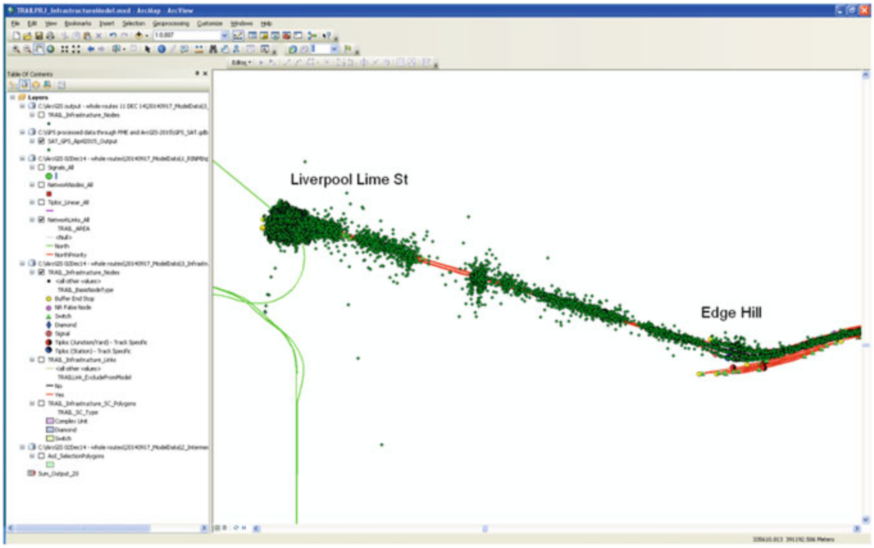


Fig. 14.14 GPS readings for multiple services at Liverpool Lime St. terminal station

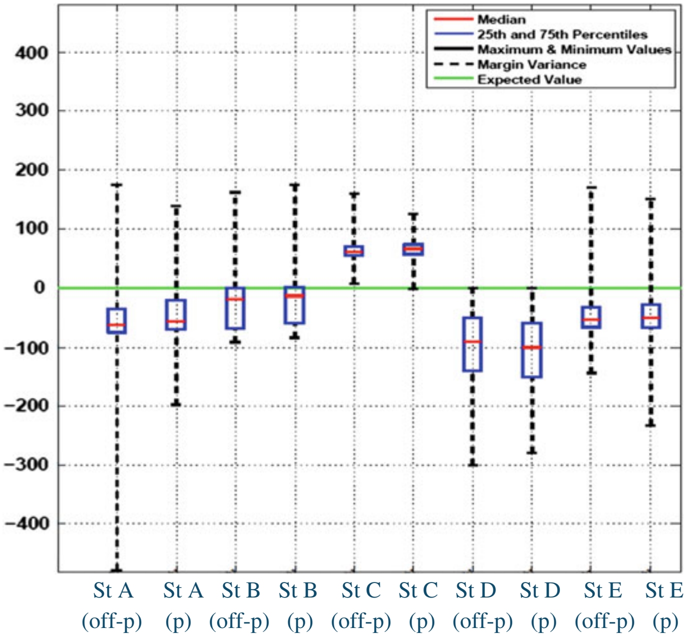


Fig. 14.15 Station dwell time variance

variance of difference between observed and planned dwell times have been plotted against the stations for peak and off-peak hours. The box covers the values that fall into 25th and 75th percentile, the red line shows the median and the whisker maps the data variance and the max and min values. These charts are specifically valuable when we need a good visual basis to compare the variables.

As shown in Fig. 14.15, the observed dwell times are mostly within the planned values apart from Station C where nearly all of the observed dwell times exceed that planned values. This means that the stopping services at Station C are affected by a frequent delay which affects the performance. This will then need to get investigated to define options to resolve this overrunning issue.

14.7.4 Observed Running Time Analysis

As also mentioned before, one of the factors that has shown to have a substantial impact on the train performance is the running time variance. The GPS data were also analysed to calculate the running times between two Timing point locations with the combination of stopping pattern as used in the timetable. An example for the observed running time calculation for two consecutive Tiplocs has been shown in Fig. 14.16. As you can see, a box and whisker has been developed for different rolling stock classes.

The variance in the observed running time is mainly due to different driving styles which has proved to have a significant impact on service performance.

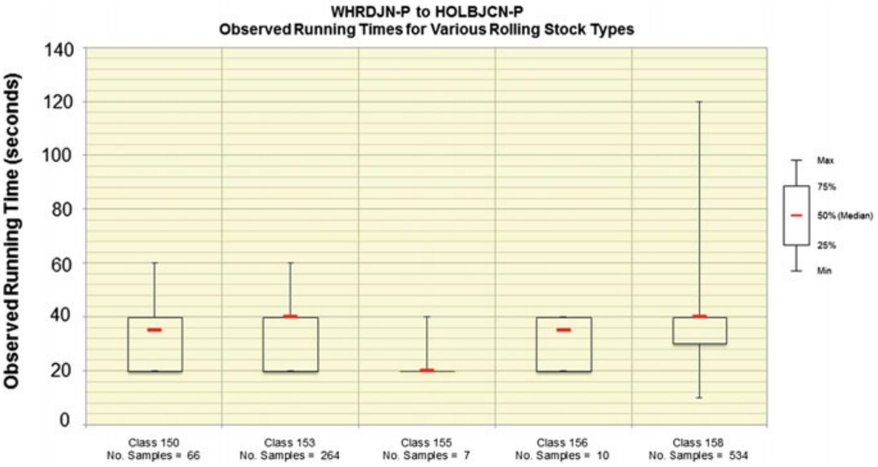


Fig. 14.16 Observed running time variance

14.8 Conclusions and Future Opportunities

The need to deliver a high performance railway in the face of increasing passenger and freight traffic is placing significant emphasis on understanding of the system behaviour and effective usage of the network capacity. The methodology developed for the study is the first time where established rail industry systems analysis processes and tools have been used together to produce a simulated representation of common cause and special cause variance to enable system performance testing, supported by both modelled and observed evidence. Also, a new series of metrics have had to be developed at a much finer granularity than was previously monitored by industry.

The results of this novel modelling and analysis technique has demonstrated that as well as improving general railway system reliability, the mitigation of Common Cause events offers a significant opportunity to deliver a more robust system with higher levels of performance. In order to target these Common Cause events, a detailed understanding of their impact on service performance is essential. Whilst infrastructure-borne performance data are readily available within Network Rail, the granularity of these data is often insufficient to accurately capture the impact of these events and the associated root causes. Based on the cases study described in this paper, train-borne measurements have offered a more detailed insight into the running time variance, station dwell time variance, and braking rates. The train-borne data also provides a rigorous basis as an evidence for the validation of journey time modelling.

Although train-borne measurements have provided a more detailed view of service performance, opportunities exist to integrate all the available data layers including:

- Signalling System State (Signal Aspect vs. Time)
- Weather Conditions (e.g. Fog affecting visibility)
- Train loading
- Level of platform crowding
- In order to deliver a rapid whole system analysis, we need to improve:
- The data acquisition processes e.g. automatic downloading of train-borne data at the end of journeys.
- The timing resolution of GPS measurements currently available
- The precision of GPS train location measurements to identify the running line
- Data format consistency and a common data language.

Finally, prejudged outcomes, biases, assumed positions can be countered through the use of clearly presented evidence derived from both observation and modelling, in short, the physics is often the only way to breakthrough pre-existing beliefs.

Network Rail's Systems Analysis Team is currently working closely with industry leaders to apply Systems Thinking methodologies to enhance our understanding of the GB railway system.

References

1. Network Rail (2015) Our history—network rail. http://www.networkrail.co.uk/aspx/729.aspx
2. Department for Transport (2007) Delivering a sustainable railway—white paper. Department for Transport. https://www.gov.uk/government/publications/delivering-a-sustainable-railway
3. RSSB (2014) Reliability modelling of the Brighton mainline (T1019)" http://www.rssb.co.uk/pages/research-catalogue/t1019.aspx
4. Network Rail (2012) CP5 strategic business plan supporting document. http://www.networkrail.co.uk/browse%20documents/strategicbusinessplan/cp5/supporting%20documents/outputs/performance%20plan%20for%20cp5.pdf?cd=2
5. RSSB (2012) Making the case for a whole system strategic approach to reliability improvement: modelling reliability study (T935 Report). http://www.sparkrail.org/Lists/Records/DispForm.aspx?ID=1507
6. Open Train Times (2015). http://www.opentraintimes.com/maps/signalling/lec1
7. RSSB (2014) Railway Group Standard GM/RT 2472. Requirements for data recorders on trains. RSSB, London, UK
8. Leeming D (2013) Right time performance measure. Ove Arup & Partners Ltd, London, UK
9. Haskins C (ed) (2007) Systems engineering handbook: a guide for system life cycle processes and activities. Version 3.1. Accessed by Forsberg K, Krueger, M. INCOSE, San Diego, CA, USA
10. Haskins C (ed) (2010) Systems engineering handbook: a guide for system life cycle processes and activities. Version 3.2. Accessed by Krueger M, Walden D, Hamelin RD. INCOSE, San Diego, CA, USA
11. Mankins JC (1995) Technology readiness levels. White paper, NASA Office of Space Access and Technology. http://www.hq.nasa.gov/office/codeq/trl/trl.pdf
12. Pollan M (2006) The Omnivore's Dilemma: a natural history of four meals. Penguin, New York, NY, USA
13. Sheard SA (2006) Definition of the Sciences of Complex Systems. INSIGHT 9(1):25
14. US Department of Defense (2003) Department of defense directive 5000.1. The defense acquisition system. Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, Washington, DC, USA
15. Deming, W. E. (1982) Out of the Crisis: Quality, Productivity and Competitive Position ISBN 0-521-30553-5
16. Shewhart WA (1931) Economic control of quality of manufactured product. D. Van Nostrand Company, Inc., New York City, p 7. OCLC 1045408
17. Thomson J (2012) Common-mode failure considerations in high-integrity C&I systems (PDF). Safety Eng
18. Calvano CN, John P (2004) Systems engineering in an age of complexity. www.calhoun.nps.edu
19. RSSB (2014) Reliability modelling of the brighton mainline (T1019). http://www.rssb.co.uk/pages/research-catalogue/t1019.aspx
20. RSSB (2016, expected published date) Brighton mainline timetable optimisation study (T1063). http://www.rssb.co.uk/pages/research-catalogue/t1063.aspx
21. Network Rail (2015) Our plan for rail in Britain—network rail. http://www.networkrail.co.uk/improvements/indexpage.aspx
22. Network Rail (2015) Performance—network rail. http://www.networkrail.co.uk/about/performance/
23. Deming WE (1982) Out of the crisis: quality, productivity and competitive position. ISBN: 0-521-30553-5

Chapter 15
Warranty Impacts from No Fault Found (NFF) and an Impact Avoidance Benchmarking Tool

Piotr Sydor, Rohit Kavade and Christopher J. Hockley

Abstract In the automotive industry the occurrence of No Fault Found (NFF) events is considered to be one of the major threats to the overall reliability and customer satisfaction. It has become essential for automotive manufacturers to carry out quick and effective fault diagnostics to identify the root cause of faults so as to avoid NFF events. Automotive manufacturers need to reduce NFF so as to reduce warranty costs as it has been recognized as one of the most significant costs in their industry and has a major impact on customer satisfaction and their profitability. Research work in the aerospace industry has developed a NFF benchmarking tool designed to address the identification of where NFF costs can be reduced through process, procedural and cultural changes. NFF in the aerospace industry though is particularly concerned with costs of day to day operation and warranty issues are less of an issue. NFF events in a high volume industry such as automotive take on a different character whereby customer satisfaction and brand success is critical to profitability. Using an adapted benchmarking tool in an industry concerned with costs generated by warranty claims will identify the non-value added activities in the fault diagnostic process and provide mitigation strategies to address NFF. The chapter will describe the differences in the impact of NFF to aerospace where warranty costs resulting from NFF are less critical to the impact in an industry where warranty and customer satisfaction with new products is critical to its profitability and success.

15.1 Introduction

The problem of No Fault Found (NFF) is a widely known and recognised especially among those who deal with complex systems.

P. Sydor (✉) · R. Kavade · C.J. Hockley
EPSRC Centre for Innovative Manufacturing in Through-life Engineering Services,
Cranfield University, Cranfield, UK
e-mail: p.sydor@cranfield.ac.uk

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_15

15.1.1 No Fault Found—Background

Following study of the literature, [1, 2], the most common set of maintenance activities resulting in the NFF can be described as follows:

> When a unit is diagnosed as faulty at the "on-platform" test level, it is removed from the system and sent to its subsequent "off-platform" test level for further investigation. If the recognition or localisation of the fault is unsuccessful then the unit in question is tagged as NFF.

Thus the often used generic definition of a fault which is classified as NFF is "a reported fault for which the root cause cannot be found" [3]. Such faults and the way they affect systems are also described in literature as: erroneous removal (ER), no problem found (NPF), can-not duplicate (CND), and re-test OK (RTOK). Attempts to provide a systematic approach to classify and describe the NFF phenomena can be found in literature [4, 5]. However, despite great efforts from the researchers and industrial practitioners, the problem of NFF still remains an open and challenging area.

A simplified example of a maintenance process is shown in Fig. 15.1. It can be seen that the NFF issue can occur at a various steps within the maintenance process. The sequence of maintenance, from fault reporting to fault resolution, can be described as follows: *A system operator records an error (e.g. fault code), maintenance personnel are notified to find the cause and provide a solution to the*

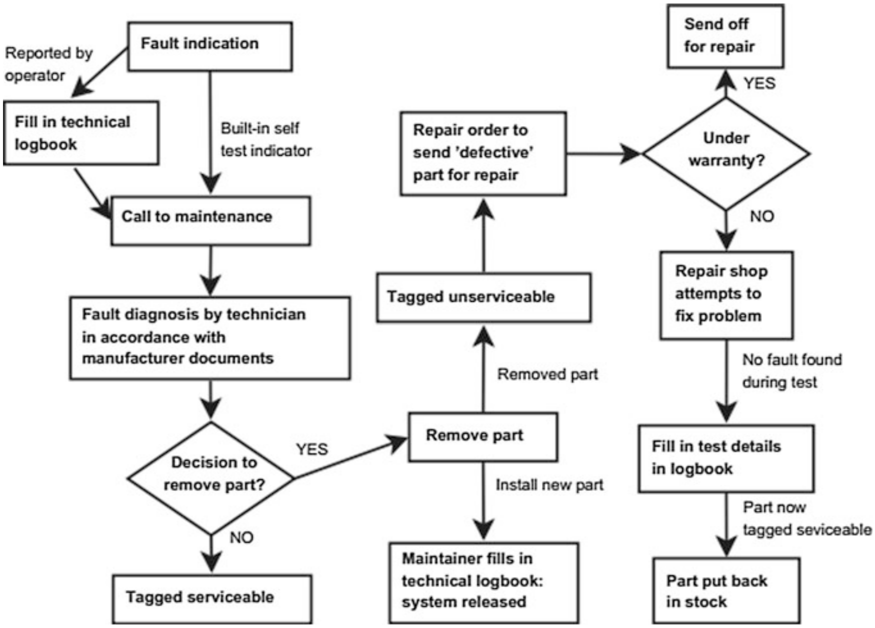


Fig. 15.1 Repair process during a maintenance action [4]

reported problem. If no root causes can be determined, the reported problem will be tagged as 'NFF'.

There may be one or many reasons that contribute to such an outcome, and it is described in more detail in this chapter when NFF root causes are discussed. In literature such situations are often associated with faults concerning electric and electronic systems, both in aerospace [6] as well as in the automotive [7] application.

15.1.2 Common Causes of No Fault Found

The following categories of NFF causes are indicated in available literature about NFF: the utilisation of a system and the support or maintenance of a system. Within these categories, there are many potential direct and indirect causes of NFF and to list them all would be impractical. Attempts to classify and provide a taxonomy of NFF, however, can be seen in various sources. Khan and Phillips [8] divided the potential causes and grouped them into the following four categories:

1. Fault diagnosis
2. System design
3. Human factors
4. Data management.

Fault diagnostics will include a range of technical equipment together with diagnostic processes and procedures some of which will use specialist test and diagnostic equipment. Diagnostic equipment also includes sensors and Built-in Test (BIT) routines or processes and the hardware that operates these tests, the BIT equipment (BITE). Processes and procedures will also include manuals, sometimes called Fault Isolation Manuals (FIMs). These manuals will often specify calibration standards for the equipment or for the process itself and include environmental considerations if appropriate. When diagnostic processes and procedures are inadequate, inappropriate, or the thresholds for pass or fail have been incorrectly set, then the diagnostic process will result in a NFF. If the sensors or the BITE are faulty or operate intermittently, then again the diagnostic process may result in NFF. If the BIT process or procedure has been poorly designed with inappropriate thresholds then NFF will inevitably result. This situation leads us onto the second cause, system design.

System design includes not only the design of the test and fault diagnostic procedure but also the design and clarity of the manuals and publications that are provided. These should always be improved and developed using feedback from operational circumstances that bring realistic diagnostic experience into continually improving the diagnostic processes. If this feedback, doesn't happen then design of the processes will still rely on theoretical information which may be wide of the mark. Design of diagnostic equipment and processes will also initially be based on a theoretical foundation of knowledge so will need to be improved using actual operational experience to ensure it does not generate NFF.

The third category to cause NFF is Human factors. There is often a reluctance to highlight the inability to find a fault where the engineer does not want to appear unsuccessful. Other human factors such as pressure to deliver a result and being seen to do something may result in the wrong component being changed. In this situation NFF may not be declared, but was in fact the real result. Lack of time, taking short cuts, lack of training and experience can all be contributory factors in causing a NFF. Pressure due to warranty claims will also feature in some industries and can be loosely grouped under the Human Factor causes of NFF. Both lack of communication and misinterpretation of symptoms are also examples where NFF can easily be caused or be the result.

The final category is data management which is a big area and particularly focuses on the ability to have enough data and data of the right quality available for fault resolution. It stems right from the time when the fault is reported; if this reporting is inadequate, the data provided can lead the diagnostic technician down the wrong diagnostic path and a NFF will occur. If knowledge of conditions when the fault occurred is not available the maintenance technician will not have a comprehensive set of data available with which to start the diagnosis. Good clear data is essential in first describing the fault symptoms and the operational conditions when the fault first occurred. Ensuring that all the relevant data is available is therefore critical as is past fault history so that data or fault trending can be investigated. Data mining in a comprehensive and accurate database is a key aspect of reducing the occurrence of NFF.

15.2 Cost of No Fault Found

The real cost of NFF remains unclear, as many studies have shown. However, the cost, dependability and safety, are listed as a critical stakeholder requirements which may be affected by NFF events [1].

A method to estimate a true cost of NFF in military systems was given in a presentation to the Machine failure Prevention Techniques Conference in 2014 and 2015 [9, 10] and the subsequent report [11]. In the work, a distinction is made between the costs of platform unavailability due to NFF issues and other maintenance and support actions. Another example of quantifying the cost associated with NFF events has been proposed in [12] where the described research outcomes provides stakeholders with a method to estimate the cost of NFF within their organisation and also across the whole supply chain. The proposed methodology in [12] is to estimate the cost of NFF events has been based on the Soft Systems Methodology proposed initially by Checkland in 1989 [13]. The key is to capture the main cost drivers of NFF across the supply chain and to build a framework to estimate the cost associated with NFF events. The Checkland's Soft Systems Methodology has been adopted and a three-phase method for NFF cost estimation has been offered in [12]. In the *Phase 1* the scope of the problem is defined, the Phase 2 deals with defining the data analysis process, and the Phase 3 focuses on the main cost drivers analysis as defined in the previous stages. The proposed method is in essence a dynamic time based modelling approach that is applied to represent the cost of NFF across the supply chain. The NFF events cost estimation model has been developed using an agent based modelling approach. Whiting the model two main groups of agents are included: the corrective maintenance agents and preventive maintenance agents. Each group consists of three agents to represent the supply chain, namely: Customer, original equipment manufacturer (OEM) and supply chain. Input to this model requires the cost data, i.e. preventive and corrective maintenance cost, as well as weighted coefficients related to each cost driver determined in *Phase 2*. Based on the above, the proposed methodology enables to estimate the total cost of NFF events. Predominantly, it is a design towards identification of abnormal cost drivers and its behaviour and evaluation of associated performance metrics with cost implications to allow analytical and heuristic information sources to be used alongside process history, costs and associated risks. The key drivers in both approaches is similar—the knowledge of how an NFF affects the overall system in terms of costs helps to determine where to concentrate and where to minimise effort in tackling the NFF root causes.

Different organisations and stakeholders will assess or count the cost in different ways. Operators and users will see the cost in loss of availability, providers of support services will see NFF as a cost in maintenance man hours for faults that cannot be found or that are repeated without resolution. Different businesses will in fact have many different and perhaps specific sources of costs that need to be identified. Once they have been there will be an assessment of the impact of these costs within whatever contract arrangements are in force.

For Operations Departments, NFF will generally reduce availability while maintenance staff struggle to find a reported fault and perhaps end up changing several items to ensure that the fault is removed, so ensuring one of the items will be tested NFF further down the supply chain. There will be system or machine unavailability and this may affect reputation and future business. Operators will also see a cost in warranty cover with equipment being returned for faults than cannot be replicated. Depending on contractual arrangements and the industry concerned the cost will fall on the operator, the distributor or the OEM. Maintenance providers will see the cost as lost man-hours and lengthy diagnostic investigations where there is no clear resolution of the reported fault; this may involve re-testing which could be expensive in the use of facilities and resources. Stakeholders in general may also have a safety cost and this can difficult to assess in merely financial terms. OEMs will have costs of re-design and almost certainly warranty costs where an endemic fault is difficult to resolve but must be solved to protect reputation and future business. There are also costs all the way down the supply chain where suppliers need to protect their own reputations and will have costs for re-testing or supply of additional parts.

There will also be indirect costs throughout the supply chain where disruption and certainty occur because of a NFF situation. Uncertainty as to whether the fault has been resolved can also feed back to the safety considerations and the need for additional checks and supervision in order to clear the system for further operation when no fault can be reproduced. The uncertainty will require further checks and balances and supervisory effort, particularly where there is no redundancy or for safety critical systems.

Finally loss of profit will occur in some industries and organisations and often needs to be balanced perhaps with considerations of safety or reputation. In some cases it is a perfectly valid option to accept the costs of NFF further down the supply chain if loss of reputation would occur. Customer satisfaction may well be more important and any detrimental effect from faults taking time to be resolved would affect business profitability in the future. Each industry and organisation must first identify the costs of dealing with NFF in order to be able to balance the decisions of where these costs should best lie.

15.3 Impact of No Fault Found

Ongoing efforts to estimate the real impact of NFF events can be seen across different industries and in the published academic research work. It has been shown already that NFF is often being misreported and is thus under-estimated, hence there is a great difficulty in assessing the true impact and cost associated with it. The following four areas of an equipment life-cycle and operation life are mostly impacted: operation and maintenance, stakeholders, original equipment manufacturers (OEM) and the overarching supply chain [12]. When operations and maintenance are considered, the NFF impacts can be seen in the lost man hours, direct maintenance cost, warranty cover, production cost, machine unavailability as well as further intangible costs, such as the loss of future business. This often coincides with the impact on the stakeholders which can be observed in the increasing costs and losses to the business as a result of getting a bad reputation, warranty cover cost, cost of in-tolerance failures and system operation training and safety. The NFF impact can be observed throughout the entire value chain. OEMs are mostly impacted by the direct and indirect capital expenditure cost such as those costs due to increased inventory maintenance, warranty liabilities, obsolescence and repair cost. Further to that the impact on the overall supply chain, with a strong emphasis on reverse logistics, can be observed by the increasing intangible cost for example due to loss in productivity, packaging and handling costs, a downtime and transportation cost.

An important fact is that beyond the direct impact on the maintenance, such as the costs of components and manpower, which are relatively easy to quantify once the NFF is recognised, there are other major impacts, often hidden, that are not easily quantifiably and its mechanisms are not fully understood yet. An example can be the loss of customer confidence and company reputation [3].

Regardless whether it is a military or a civil system, the occurrence of NFF events causes a disruptive effect on the successful delivery of through-life customer support. Whilst commercially NFF may cause a huge loss of revenue, in the military environment, where costs are less obvious and may be hidden, losses may be even higher. A lack of visibility of the true cost is an issue that must be tackled in the most informed manner. It is essential to stress here that numerous maintenance terms are used rather than declare a 'no defect' situation; terms such as CND and RTOK are all NFF and are important contributors to the overall hidden costs that nugatory maintenance causes which in turn deliver potentially high levels of operational disruption.

A significant impact due to NFF events can be also registered in the high cost of warranty returns. If a product is warranted and returns are much larger than initially forecasted it may generate high cost to the manufacturer, both in the fiscal terms as well as the reputation for unreliability and high rate product replacements.

15.3.1 NFF in Aerospace Industry

The NFF issues are certainly not restricted to any particular industry, however it can be observed that the aircraft industry has been the strongest advocate of tackling the NFFs as equipment is more expensive and the NFF incurred downtime causes significant loss in revenue. A nugatory and wasteful maintenance activity generates high losses and greatly contributes to the disruption to the overall operation. In-service support, that includes maintenance and repair, constitute the majority of life-cycle of aero platforms. It has been argued in the literature that for equipment having an in service life of around 20 years, such as aero-platforms, the operating and service/maintenance activities may account for as high as about a 60–80% of the whole life cycle cost of the equipment [11, 12].

In the case of aircraft systems, NFF events manifest themselves during a high stress, such as high g-forces, extreme thermal cycles, high vibration levels, or a combination of stresses [11]. High percentage of aircraft NFF problems, mostly accredited to the intermittent faults, are related to aircraft ageing, especially to problems with wiring and connections as a result of environmental and operational conditions such as exposure to vibration resulting in cracking, oxidation, heat cycling and spark-erosion, etc. [14]. It is a growing problem, mostly in legacy aircraft systems, where the effects on electric and electronic equipment, such as broken wires, cracked solder joints or corroded wire wrap are more prevalent [1].

Maintenance, repair and overhaul activities are also recognised to significantly contribute to the NFF rate [4, 10]. An often encountered inability to recreate a fault condition during bench testing results in the component being declared NFF or CND and returned as serviceable. The issue here is that bench testing is either not representative of the flight conditions or has higher thresholds for failure. The test bench itself may not be sufficiently sensitive to find faults and to determine the root cause of the original fault or symptoms seen on the aircraft. Quite often faults that have no relevance to the original fault will be identified on the test bench and consequently repaired yet these had no relevance to the original fault which will now remain dormant. Whilst not recorded as NFF on this occasion, the fact that the real cause of the fault has not been found will ensure that the root cause still becomes a driver of future fault symptoms and removals which will be designated NFF.

15.3.2 NFF in Automotive Industry

The automotive industry has been and still is evolving at a very high pace. This industry is seen as a leading exponent of setting new standards and trends within various engineering domains. This perhaps makes it more vulnerable to NFF problems, when compared with aerospace industry. The cost of single NFF incident in an automotive system is considerably lower than that in an aerospace system, especially when considering the potential severity and the direct and indirect effect it has on its ecosystem. To demonstrate this, an example can be considered from an everyday operational scenario; an immobilised road vehicle can case a minor disruption on the road and in most cases a low cost of recovery, especially when compared with the ripple effect that an immobilised air vehicle can cause on the runway or within a busy airport and combined with the high cost of recovery and subsequent maintenance costs.

Production demands in the automotive industry are ever increasing. According to the Society of Motor Manufacturers and Traders, in 2014 the production of passenger cars in the UK was around 1.53 million units with Nissan, Land Rover, and MINI being the leading brands contributing to these production figures. Future demand in this sector is also expected to increase as car manufacturing volumes are on course to break all-time records by 2018. With the increase in demand, customers of the automotive industry are also showing strong interest for more in-vehicle technologies and digital services. Just as phones got smart, so will cars. The rise in the digital economy and an increase in demand for customised products have caused modern automotive vehicles to become a complex system [15]. Cars are now well equipped with ever more complex systems that are connected with other remote systems and data centres, e.g. warning devices, navigation and traffic information services, infotainment systems and safety features, and thus are commonly called networked cars. The number of networked cars will rise by 30% a year for the next few years. It is estimated that by 2020, one in five cars will be connected to the Internet [15]. Thus the usage of embedded software and electronic systems will play a dominant role in the coming years for the automotive industry. The OEMs are therefore looking for ever more innovative solutions; products with shorter time to market will also be required in order to attract customers and maintain market share. Figure 15.2 shows the current usage of electronic systems on a normal passenger car [16].

When a unit or component is diagnosed as faulty at any test level, the unit is removed from the system and sent for test at the next maintenance level. If the diagnosis or localisation of the fault is unsuccessful then their unit is simply tagged

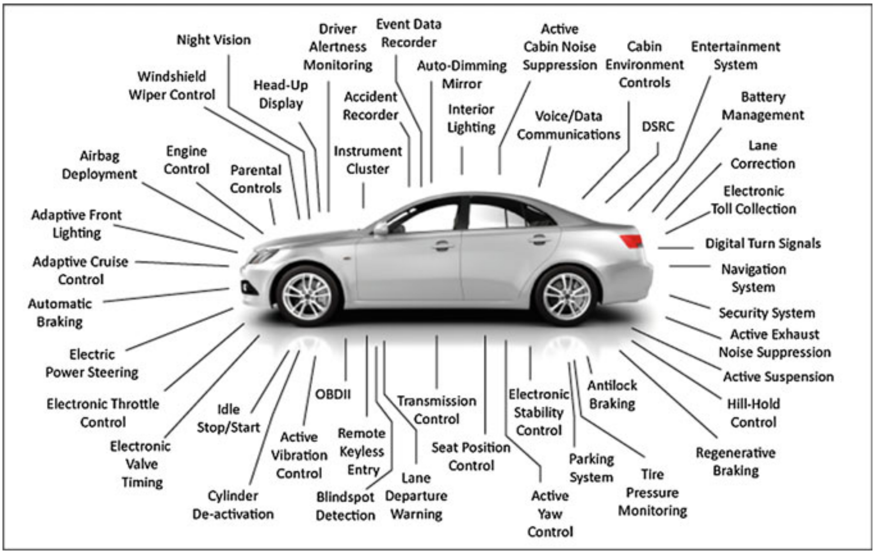


Fig. 15.2 Automotive electronics [16]

as NFF [1]. A common definition of NFF, also to be found in [3], is "*a reported fault for which root cause cannot be found*". NFF and intermittent failures have been reported in the automotive, aerospace, telecommunications, computer and consumer industries where they represent a significant percentage of reported warranty returns and field returns, resulting in significant costs [17]. The phenomenon of NFF is very well known in the aerospace industry, and particularly in avionics. A study performed by Boeing, Texas Instrument and General Dynamics and reported by Pecht [18] showed various military and commercial avionics systems and radar systems which had been analysed, showed that, depending on the type of system, between 21 and 70% of cases have been attributed to NFF. Also in more recent studies [19] similar rates of NFF can be observed in automotive, telecommunications, computers and consumer electronics industries. However, in the automotive industry it is less understood as a problem for availability, rather it is a problem for the distributors and the OEMs of warranty and maintaining reputation. Consequently the reliability of the complex electronic systems has become an important issue for the automotive industry.

The architecture of an electronics system in an automobile includes a combination of different sensors, electronic control units (ECUs), actuators and user interfaces to perform complex electrical functions [20]. Due to the addition of flourishing on-vehicle services, user and software functions are interrelated at the component level and at the whole vehicle network level. Communication between the components has become difficult to manage and performance is hard to predict in terms of the run-time behaviour of these networks [21]. The complexity brought about by embedded software and electronics has created unavoidable challenges in maintenance and repair, threatening customer satisfaction and causing negative effects on the costs of repair and replacements. From the software engineering perspective, around one third of problems are associated with software and electronics [15]. This high level of complexity and associated failures in vehicle electronics has given rise to a significant increase in NFF events and warrant issues for many manufacturers.

The NFF phenomenon also can have a negative impact on the system safety, dependability and it inevitably increases lifecycle costs. The manufacturer, component and system suppliers, the service centres and of course the operators, or users, are all considered as stakeholders of NFF. NFF events can be basically classified into two categories: those that affect safety and those that do not. One sub-category of NFF events that affect safety is where a test process at the user or operator level fails to recognize and localize the reported fault. Another safety-related sub-category is NFF events where the subsequent test process at the next service and repair level is unable to localize the actual fault. However, an NFF event where at the preceding test level, a fault is recognized that does not exist, i.e. a false alarm, this does not directly affect safety. Unfortunately, it is not always easy to decide which category of NFF event has actually occurred [22].

The major impact of NFF in the automotive sector is on the warranty or support budget which is wasted trying to find the root cause of the reported faults. Such events tend to increase the spare parts inventory and the cost of work and manpower [7]. Very often reported faults are due to operator or user error and their unfamiliarity in using and operating the system. Within the aerospace industry NFF can cause up to 90% of the total maintenance costs related to aircraft electronics. The automotive industry, like many other manufacturing industries, struggles with the high cost of improving fault diagnostics. Some authorities claim that the cost of poor fault diagnosis can be as high as 20–30% of finished products [1]. Consequently it is a big issue that no manufacturing company can afford to ignore.

Without understanding the root cause, companies and particularly the automotive dealerships, often find themselves implementing a workaround or swap-out solution of the suspect part with no real idea of whether it will work. Considering the major impacts of NFF issues, any step-change improvements are scarce and hence the lack of data on the cost impact means that dealing with the problem is unlikely to be escalated. Since "what gets measured gets done", there is consequently a need for a generic tool that can be used (within any industry) to evaluate an organization's ability to deal with the problem.

15.4 Solution

Once an organisation has realised that there is a NFF problem, a useful solution is to use a NFF Impact Avoidance Benchmarking Tool. Such a tool was developed by the EPSRC Innovative Manufacturing Centre for Through-Life Engineering Services at Cranfield University (The TES Centre) and has been applied in both aerospace and the automotive industry. The tool enables a thorough analysis of the NFF problem and the processes used in the organisations together with identifying the attitudes and culture surrounding its resolution. The tool was developed after extensive research and industry consultation allowing key aspects of the NFF process of where NFF occurs to be established. The breakdown of the processes was important for each different organisation in order to identify and establish where root causes were identified and where faults were diagnosed successfully. However, there are also difficult organizational concerns and human factors aspects to take into account as they also cause NFF in addition to technical causes; such organizational, human factor and process issues are usually easier solve and to provide mitigations for once they can be identified. The tool thus concentrates the mind on the process and the underlying attitudes using a team approach where a number of engaged and knowledgeable experts assess the ability to reduce NFF.

Benchmarking using this tool approach is a structured way of assessing an organisation and where changes are required and may they may be accepted and adopted. Benchmarking also identifies areas where easy wins can be achieved and thus contributes to a continuous improvement process (CIP) for the organization. The main purpose of the benchmarking tool is defined as [10]:

> "To achieve a minimum level of non-value added activity in the timely diagnosis and resolution of complex system fault indications with the minimum amount of time and other resources expended to confidently restore the customers operation, robustly isolate and remove the cause of the fault indication, and to learn lessons and prevent future impacts."

The process model created was initially based on the aerospace industry and includes the on-platform and off-platform processes which may or may not result in NFF. Customization of the process to adapt to the automotive industry was simply done using inputs from dealership, OEM and distributors. A high-level view of the generic maintenance process in which NFF occurs is shown in Fig. 15.3.

The first step is thus to adapt the model with any changes to the process. Once agreed the assembled team of subject matter and departmental experts, uses it as a self-assessment tool which requires some subjective assessment in specifying the target performance level for each process step. The tool then scores the selections and the team scores can be compared or averaged. Differences in scoring are not important and indeed demonstrate different perceptions across the organisation. Any low score allow the team to highlight where the organization is weak and enables it to focus on solutions and options for mitigation.

Whilst it was initially developed for the aerospace industry, it has been easily adapted and tailored to suit the automotive industry, thus proving its versatility. The key motivations across all industries are universal—to uncover where NFF is impacting maintenance and warranty costs and to provide mitigations, recommendations and solutions. The key steps of the benchmarking tool which should be taken into account when adoption into different class of problems are shown in Fig. 15.4.

The possible mitigation strategies were developed from expert knowledge, best practice examples and from the accumulated research knowledge within the

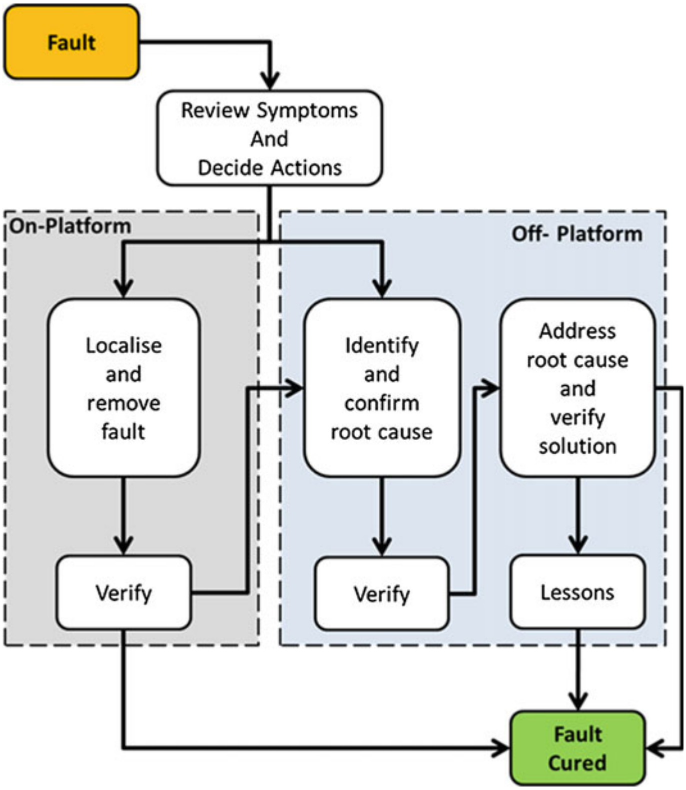
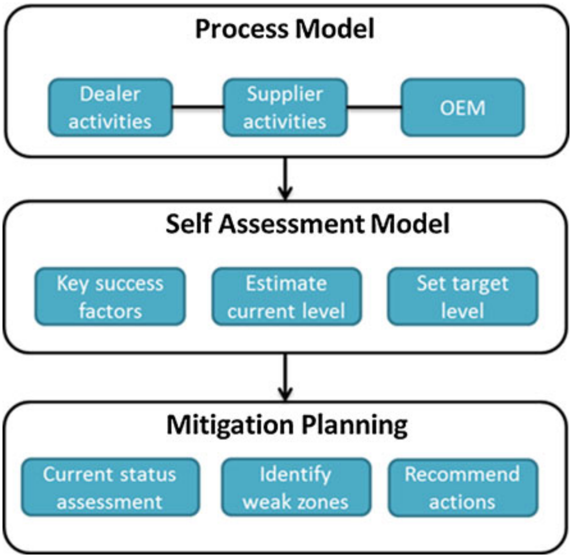


Fig. 15.3 Generic model of maintenance action

Cranfield TES Centre. Potential actions for end-user organizations include for example, new tools, where support is needed, required infrastructure improvements, team actions and training; mitigation strategies will always need to be customized but are provided to allow an organization to move its current assessed capability and performance, to the desired higher level in avoiding and reducing NFF issues. The tool allows mitigation options to be pre-set and to be generated as a direct result of feedback and post-analysis. The organization can decide then to make any necessary transition slowly (evolutionary), or take a more rapid approach (revolutionary). The organization's own experts will be the best judge of what is possible and can best be delivered by revolution or evolution.

In order to address the NFF issues specific to the automotive industry an NFF impact avoidance benchmarking tool has been developed that will analyse the ability of automotive suppliers and manufacturers to deal with the NFF problems. The tool was based on the previous research work that developed a 'NFF benchmarking tool for the aerospace industry' The approach has been evaluated and tested in partnership with an automotive manufacturer. The combined knowledge—gathered from the literature survey on NFF, the background research of the

Fig. 15.4 Main steps of the NFF impact avoidance benchmarking process



automotive industry as well as the internal procedures of the partner organisation helped to define the problem and to deliver a suitable solution. The key objective was to control the warranty cost attributed to NFF, and to achieve a minimum level of non-value added activity in the timely diagnosis and resolution of complex system fault indications with the minimum amount of time and other resources expended.

15.5 Conclusions

NFF is a serious problem across many industries yet many organisations do address the problem and this can be for a variety of reasons. Most common is the inability to assess the true costs and whilst the impact may be noticed, there is an inability to identify mitigation actions that would reduce the problem. Cost and impact are difficult to estimate without good information. Some examples of cost and impact modelling are available but useful literature and descriptions are few and far between. Addressing the issue by using a benchmarking tool such as described in this chapter will be an essential first step in identifying the process around NFF occurrence and will systematically expose where to concentrate effort for the most rapid gain and progress.

The impact of NFF in aerospace industry mostly relates to high value components and its effects, such as disruption to service and safety breaches, whereas in the automotive industry, the key impacts are felt with much higher volume of albeit, less-costly effects. In both cases, however, the stakes are high.

In the automotive industry the occurrence of NFF events is considered to be one of the major threats to the overall reliability of the system and an important factor in warranty costs. Following the example of aerospace industry, it has become essential for the automotive manufacturers to avoid NFF events by means of effective fault diagnostics and correct identification of root causes of fault at every step of the maintenance process. This will certainly help minimise the impact of NFF, and subsequently help to lower the cost of warranty claims. However, communication between the dealerships and the OEMs is poor and yet to be properly incentivised so that both share and solve the problem.

An excellent proposed solution, as a first step, is to identify the non-value added activities in the fault diagnostic process and provide mitigation strategies to address the NFF events as proposed in the NFF Benchmarking method described in this Chapter.

References

1. Söderholm P (2007) A system view of the no fault found (NFF) phenomenon. Reliab Eng & Syst Saf 92(1):1–14
2. Khan S, Phillips P, Hockley CJ, Jennions IK (2012) No fault found, retest-ok, cannot duplicate or fault not found?—towards a standardised taxonomy. In: Roy R, Shehab E, Hockley CJ, Khan S (eds) Proceedings of the 1st international through-life engineering services conference, EPSRC centre for innovative manufacturing through-life engineering services, Cranfield University Press, Cranfield Universtiy, Cranfield, UK, pp 246–253
3. Cockram J, Huby G (2009) No fault found (nff) occurrences and intermittent faults: improving availability of aerospace platforms/systems by refining maintenance practices, systems of work and testing regimes to effectively identify their root causes. In: Proceedings of CEAS European air and space conference
4. Khan S, Phillips P, Hockley C, Jennions I (2014) No fault found events part 1: procedural, cultural and commercial practices. Reliab Eng Syst Saf 123(183):195
5. Khan S, Phillips P, Hockley C, Jennions I (2014) No fault found events part 2: technical root causes, testing and diagnostics. Reliab Eng Syst Saf 123:196–208
6. Qi H, Ganesan S, Pecht M (2008) No-fault-found and intermittent failures in electronic products. Microelectron Reliab 48(5):663–674
7. Thomas DA, Ayers K, Pecht M (2002) The “Trouble Not identified” phenomenon in automotive electronics. Microelectron Reliab 42(4):641–651
8. Khan S, Phillips P (2012) Tackling no fault found in maintenance engineering. In: Proceedings of the 1st annual symposium in no fault found, Shrivenham, Cranfield University, UK
9. Hockley CJ, Phillips P, Lacey L (2014) Identifying the true cost of no fault found (NFF) within the united kingdom ministry of defence (mod). In: Proceedings of the machine failure prevention technology conference
10. Hockley C, Sydor P (2015) The hidden delta for maintenance effectiveness. In: Joint conference of MFPT 2015 and ISAs 61st international instrumentation symposium, Huntsville, Alabama, USA
11. Hockley CJ, Lacey L, Pelham JG (2015) Report for air command on the impact of no fault found (NFF) on a selection of RAF aircraft fleets. Technical report TES 02-01-2015, EPSRC through-life engineering services centre, Cranfield University, UK

12. Erkoyuncu JA, Khan S, Hussain SMF, Roy R (2016) A framework to estimate the cost of no-fault found events. Int J Prod Econ 173:207–222
13. Checkland PB (1989) Soft systems methodology. Hum Syst Manage 8(4):273–289
14. Steadman B, Sievert S, Sorensen B, Berghout F (2005) Attacking "bad actor" and "no fault found" electronic boxes. In: Proceedings of IEEE Autotestcon Conference, pp 821–824
15. Mohr D, Muller N, Krieg A, Gao P, Kaas HW, Kreiger A, Hensley R (2013) The road to 2020 and beyond: whats driving the global automotive industry? Tech rep, McKinsey & Company, Inc, New York, USA
16. CVEL (2016) Automotive electronic systems, the clemson university vehicular electronics laboratory. http://www.cvel.clemson.edu/auto/systems/auto-systems.html
17. Chan HA (ed) (2001) Accelerated stress testing handbook: guide for achieving quality products. Wiley-IEEE Press
18. Pecht M, Ramappan V (1992) Are components still the major problem: a review of electronic system and device field failure returns. IEEE Trans CHMT 15(6):1160–1164
19. Bajenescu TI, Bâzu MI (2011) Failure analysis: a practical guide for manufacturers of electronic components and systems, Chap 3. Springer, Berlin Heidelberg GmbH, pp 37–70
20. Furukawa Y, Kawamura S (2006) Automotive electronics system, software, and local area network. In: Bergamaschi RA, Choi K (eds) Proceedings of the 4th international conference on hardware/software codesign and system synthesis, CODES + ISSS 2006, Seoul, Korea, 22–25 October 2006, ACM, p 2
21. Borberg B (2006) Embedded systems lifecycle management. In: Proceedings of the IBM EE engineering CAR-symposium
22. Steadman B, Pombo T, Madison I, Shively J, Kirkland L (2002) Reducing no fault found using statistical processing and an expert system. In: 2002 IEEE AUTOTESTCON proceedings on systems readiness technology conference, IEEE, Huntsville, Alabama, USA, pp 872–878

Chapter 16
Insights into the Maintenance Test Effectiveness

John Thompson and Laura Lacey

Abstract The condition of no fault found (NFF) has been accepted by engineers and operators for many years. Whilst there are known general reasons for NFF one of the areas that is overlooked is the importance of the test equipment, whether built-in test equipment or automatic test equipment, and specifically the effectiveness of the test. How much of a system or component is actually being test by the test equipment? This paper explores the system or component fault coverage effectiveness and the impact on NFF. In doing so a module is proposed to establish the relationship between NFF and test coverage effectiveness for different reliabilities. As the component or system reliability improves along with the test effectiveness the percentage of NFF occurrences decreases. Another consideration is the assumed number of random failures that the design engineers include during the development phases. This is usually an underestimate and the combination of this and the ineffective fault coverage have a negative impact on the operational availability.

Nomenclature

E	Effectiveness fault coverage of the test
MT	Maintenance test
NFF	No fault found
R	Reliability of system/component before the test
R_E	Reliability of the tested part of the system/component before test
$R^{(1-E)}$	Reliability of the NOT tested system/component
R_L	Reliability of the whole system/component after testing

J. Thompson
England, Yeovil, UK
e-mail: John.Thompson314@BTinternet.com

L. Lacey (✉)
Centre for Defence Acquisition, Cranfield University, England, Bedford, UK
e-mail: l.lacey@cranfield.ac.uk

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_16

R_T	Reliability of system/component after test
α	Probability of a FAIL-FALSE
β	Probability of a FAIL-PASS

16.1 Introduction

Working in the UK defence industry on helicopter logistic support contracts it was observed that helicopters required daily maintenance test (MT) action to find and replace failed Line Replaceable Units (LRUs). This is usually achieved by conducting a test using Built in Test Equipment (BITE) and/or using Automatic Test Equipment (ATE) and the maintainers' knowledge and practices. However, with all of these No Fault Founds (NFF) still occurs.

There is the assumption, based upon experience, that these maintenance techniques are 100% effective [E] at detecting and replacing the failed component/LRU. Moreover, it is also incorrectly assumed that the BITE and ATE test (from this point forward referred to as Test Equipment (TE)) covers 100% of the component/system and thus will identify all failure modes.

It is suggested that this is not the case and that the MT have errors such as NFF and can leave a failed component/LRU on the system. This chapter is to suggest a model to give insights into the MT effectiveness and errors and its effect on the component/system reliability from before the test (R) to after the test (R_L) and NFFs. The effectiveness of the maintenance test is not usually considered [1], neither is the condition of the component before the test. So often the component is disturbed before testing the whole system.

16.2 System and Component Test

TE is designed to test certain parameters. This may not be a complete test of the whole component or the system. This is therefore the effectiveness of the test. For this example image an LRU and use TE to test a component, the effectiveness of the test may be only 80%. That means that 20% has not been tested. A fault may still exist in this untested 20%, as shown in Fig. 16.1.

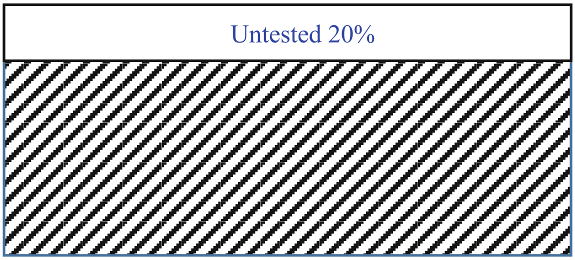


Fig. 16.1 Untested part of a component

16.3 Analysis

To develop a MT model the component/system reliability [failure rate] is partitioned into two parts. The reliability of the tested part is shown by R^E whilst the reliability of the NOT tested part is shown by $R^{(1-E)}$. So the overall reliability of the system/component is given by Eq. 16.1.

$$\boldsymbol{R} = \boldsymbol{R}^{\boldsymbol{E}}.\boldsymbol{R}^{(1-\boldsymbol{E})} \tag{16.1}$$

For the analysis it was assumed that the component was still fitted to the aircraft when the system is initially tested. A MT only has two possible initial outcomes a PASS or a FAIL and this only applies to the tested part $\left(R^E\right)$ of the component/system. These two outcomes have a further outcome:
PASS

TRUE—A correct result. The item is serviceable.
FALSE—An incorrect result. An unserviceable item has been passed. This will be identified as a failed item β probability.

FAIL

FALSE—An incorrect result. A serviceable item has been failed. This will be identified as a serviceable item with α probability. This will later be identified as a NFF.
TRUE—A correct result. The item is unserviceable.

A FAIL-FALSE item may when tested at Depth maintenance result in a NFF as the fault is not within this component.

Also, it is assumed that failures in the tested part of the component/system will be successfully repaired and that the reliability of the not tested part $\left(R^{(1-E)}\right)$ will remain the same as before the MT. Figure 16.2 shows the MT results and their probabilities.

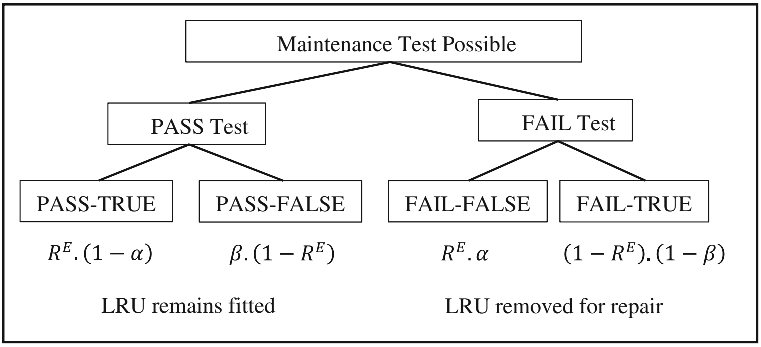


Fig. 16.2 Maintenance test possible outcomes

16.4 Discussion of the MT Model Parameters

R shall be used to indicate the reliability of the system or component before a test has been conducted. I have seen values range from 0.5 for older, in-service complex machines to 0.999999 for weapon systems. This can be assessed by a simple count of successful uses divided by the total number of uses, or by use of complex distribution models of random failures from exponential, a constant failure rate, to Weibull, using say, increasing failure rate.

$\mathbf{R^E}$ shall be used to indicate the reliability of the tested part of the component/system. As previous stated the whole system will not necessarily be tested by the test equipment.

$\boldsymbol{\alpha}$ shall be used to indicate the probability of a FAIL–FALSE (i.e. a no fault found—the component is serviceable but is removed for repair). From experience this is a subjective probability and is usually in the range 0.01–0.20 (1–20%). This includes the probability that the test equipment is too sensitive and incorrectly detects faults or the judgement of the maintainer is too harsh. This probability will be affected by the maintenance practice. If the maintainers have to ensure the system will be serviceable and available [2] and there is any doubt about the test equipment result the item may be removed. Additionally, they may replace most of the components (LRUs) in a subsystem even if this is up to four components. Thus, assuming only one LRU failed, α is now 0.75 (75%)—a high value. Alternatively, a technical manual may narrow the fault down to four LRUs and so the maintainer will remove all four LRUs just to be on the safe side. This reduces the equipment downtime but should maintain or improve the operational availability [2] so long as there are sufficient spares within the supply chain.

$\boldsymbol{\beta}$ shall be used to indicate the probability PASS–FALSE. The component has failed but remains on the aircraft. There is a fault but it has remained undetected (in the tested part) by the TE. From experience, this is a subjective probability usually in the range 0.01–0.20 (1–20%). This includes the probability that the test equipment is not sensitive enough to detect failures, so perhaps the threshold has been set too high. Again this probability is affected by the maintenance practice. If the maintainers are poorly trained and under stress they may pass the aircraft/component as serviceable when an LRU has actually failed. Questions must be asked as to whether the maintainers know how to use the test set. How frequently they use the test set? What test set refresher training is provided and at what periodicity [3]?

There is a relationship between α and β; they both cannot be above 0.5 at the same time because the total probability is greater than 1.0. Figure 16.3 below shows this relationship as a guide.

E is the effectiveness of the fault coverage of the test. It must lie somewhere between 0.0 and 1.0. This is how much of the machine/system failure rate is tested or even testable by the TE and the maintainers' knowledge. This is usually in the range of 0.5, for systems in operation and poor test equipment, to 0.99 at the original equipment manufacturer (OEM) who will normally have the best test equipment and engineers' knowledge. The maintainers' knowledge will be affected

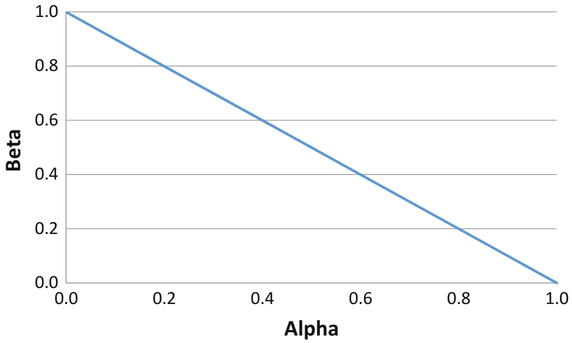


Fig. 16.3 Alpha-Beta relationship

by the training of the system and the test equipment, the manuals currency, the time available to conduct the diagnostics and the testing and final the culture of the organization [3].

An estimate of the system E is

$$\mathrm{E} = \left(1 - \left[\frac{\log(R_L)}{\log(\mathrm{R})}\right]\right) \tag{16.2}$$

where R is the reliability of the system/component before the test and R_L is the reliability of the tested part of the system. This estimate suggests the effectiveness is in the range of E = 0.50 to 0.80.

16.5 The Maintenance Test Model

There is now have a MT model to investigate the reliability of the system after the test (RL) and the quantity of NFF produced by the test. For the following explanation it is assumed that a system contains four components, namely LRUs, as depicted in Fig. 16.4 and the TE can only effectively test LRUs A and B.

So if α and β for LRUs A and B are zero, these would be serviced or repaired as necessary and it would be assumed that they are now 100% serviceable; assuming a perfect/correct repair. The reliability of the system is limited by the reliability of

Fig. 16.4 The system containing four LRUs being tested

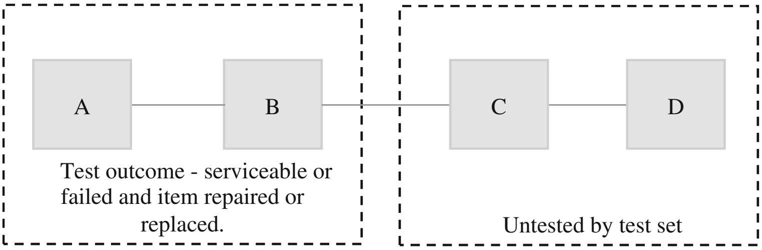


Fig. 16.5 Reliability of system

LRUs C and D, or at least the untested aspects of these LRUs. The crux of the problem is what percentage of the system does the TE test? So the reliability of the system can be shown by Fig. 16.5.

So

Firstly, the reliability after (R_L) the test:-

$$R_T = \frac{(\text{PASS} - \text{TRUE})}{[(\text{PASS} - \text{TRUE}) + (\text{PASS} - \text{FALSE})]} \tag{16.3}$$

$$R_T = \frac{R^E.(1-\alpha)}{R^E.(1-\alpha) + \beta.\left(1 - R^E\right)} \tag{16.4}$$

But, this reliability has to be multiplied by the reliability of the NOT tested part of the system/component, $R^{(1-E)}$ (See Fig. 16.1). So,

$$R_L = R_T \times R^{(1-E)} \tag{16.5}$$

$$R_L = \left(\frac{R(1-\alpha)}{R^E.(1-\alpha) + \beta.\left(1 - R^E\right)}\right) \tag{16.6}$$

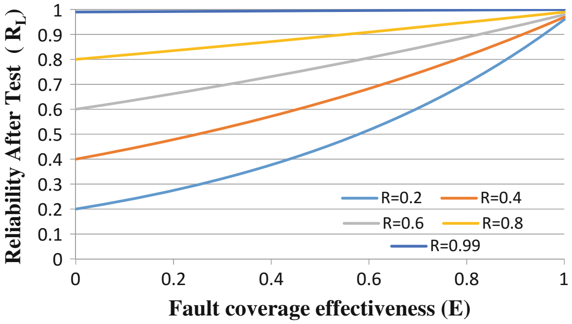
Therefore, $R_L = (R_C \text{ x } R_D)$. There is unknown confidence in LRUs C and D. If $\alpha = \beta = 0$ there are no test error terms.

Then $R_L = R^{(1-E)}$ the reliability after test (R_T) is the reliability of the NOT tested part of the component/system, when $\alpha = \beta = 0$, no test errors.

So if there have 100% fault coverage [E = 1] then $R_L = 1.0$, a perfect test and repair. If no part of the system can be tested there will be 0% fault coverage, [E = 0], then $R_L = R$. The reliability after test (R_T) is the reliability of the system/component before the test (R). The test has no effect.

Figure 16.6 is a plot of reliability after the test (R_L) against fault coverage E for various reliabilities before the test (R).

Fig. 16.6 Reliability of whole system after testing (RL) against fault coverage (E) for various reliabilities before the test (R)



16.5.1 Probability of No Fault Founds

The probability of NFF can now be determined.

$$\mathrm{NFF} = \frac{(\mathrm{FAIL}-\mathrm{FALSE})}{(\mathrm{FAIL}-\mathrm{FALSE})+(\mathrm{FAIL}-\mathrm{TRUE})} \tag{16.7}$$

$$\mathrm{NFF} = \frac{\mathrm{R}^{\mathrm{E}}.\alpha}{\left[\left(\mathrm{R}^{\mathrm{E}}.\alpha\right)+\left(1-\mathrm{R}^{\mathrm{E}}\right)(1-\beta)\right]} \tag{16.8}$$

Figure 16.7 shows the percentage NFF against MT effectiveness [E] for a range of reliabilities [R] with $\alpha = \beta = 0.05$. For the same reliabilities Fig. 16.8 shows what happens when α is decreased from 0.5 to 0.3. The percentage of NFFs increases quite substantially.

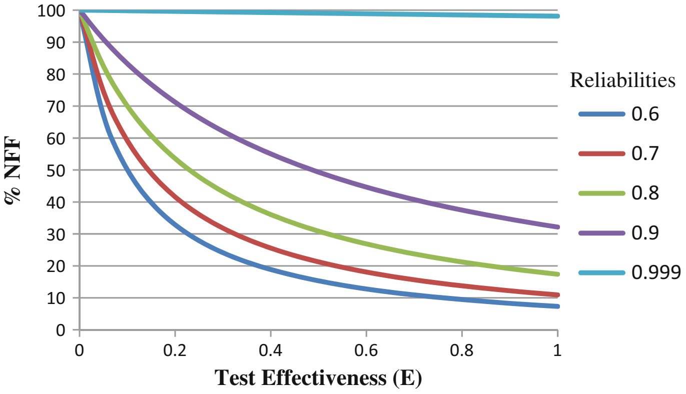


Fig. 16.7 Percentage NFF versus effectiveness for a range of reliabilities and $\alpha = \beta = 0$

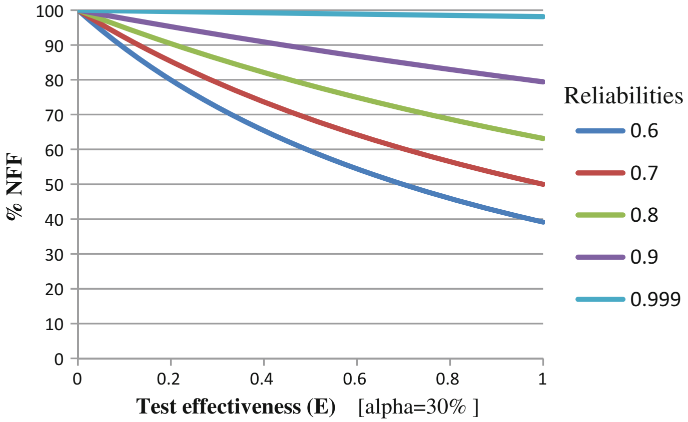


Fig. 16.8 Percentage NFF versus effectiveness for a range of reliabilities and $\alpha = 0.3$ and $\beta = 0.7$

16.6 NFF and Effectiveness

With the average MT efficiency [failure rate coverage] [E] of 80% and reliabilities between values of 0.6–0.8 the percentage of NFF is in the range 10–22%. This is a typical range of percentage for NFF that were observed in the Helicopter support contracts for military avionic components. If a component has a reliability of 0.2 but NFF is in the range of 5–10%, then should be queried. Something does not tally; either α and β could be very high or the coverage is not as expected.

When the reliability of the component/system is high, say above 0.99, the quantity of failures is reducing and there are fewer failures to detect. However, in contrast the percentage of NFF is approaching 100%. Therefore, when measuring the percentage of NFF, the reliability of the system should be taken into account. A high percentage NFF is not a major cost when the system reliability is high. This suggests that with very high reliability/system, above say 0.999, each MT will produce 100% NFF at α probability. There may be minimum time between each MT to ensure the probability of failure is high enough so the MT can detect the failures, i.e. the probability of a failure is equal to α, the probability NFF.

16.6.1 So Why Carry Out Unnecessary Tests?

The periodicity of the test needs to be queried. The question to be asked is how often does the system (e.g. a missile) need to be tested if it has a proven high reliability? Then for any frequent test there is a probability α that the result is a FAIL-FALSE. However, this is not an actual true failure but as a consequence of the test result additional maintenance is required.

16.6.2 NFF and In-Service Failure Modes Effect on Reliability [R]

A usual assumption for a component failure rate (MTBF) prediction is that the failure rate is ‘random’ with a constant hazard function and, all failure modes are attributable, i.e. no NFF. This assumes that any design or manufacturing defects have been found and corrected in the product development phase. In defense products experience by the authors has that up to an additional 40% of the ‘random’ failure rate has to be added to account for the design or manufacturing defects in the products in-service phase. It may be these may have been detected with greater Production Reliability Assessment Testing [2]. Did the prediction include a percentage of random failures and was this the same value of the actual in-service number of failures? If not it will need to be adjusted to account for reality.

16.6.3 Design, Manufacturing and Testing Error Examples

The following are some examples of design, manufacturing and testing errors that only became evident during operations.

16.6.3.1 Example—Runway Denials System

A major design failure mode is for a mine of a runway denial system, which during trials when the mine landed it detonated prematurely. The fault was traced to a gas motor that on operation generated an electromagnetic pulse, which fired the trigger circuit. The corrective action was to put a Faraday cage around the trigger circuit.

16.6.3.2 Example Military Helicopter Main Rotor Blades

Another design failure was on a military helicopter’s main-rotor blades’ leading-edge shields. There was no design requirement for the shield butt joints to be parallel and with a specified minimum clearance distance. The butt joints are not visible during operations as they are covered with a butt strap. During flights the temperature changes caused the leading edge shields, to expand and touch and buckle at the joint and thus making the butt strap loose. As a result corrective action was instigated, which was to specify a minimum distance between the two sides of the butt joints and how parallel or not they should be. Originally during manufacturing it was not possible to assemble the shields with a perfectly parallel butt joint of say 2 mm apart. So for two ‘parallel’ butt joint lengths of 50 mm original assemblies permitted a 1 mm difference between joints’ two ends, and thus 2 mm over 100 mm etc. However, with a temperature change this 1 mm gap closed and the shields touched. The design engineers had to set an off- parallel limit. For example, for a 50 mm length of shield the gap between the two butt joints had to be 0.5 mm in 50 mm but with a minimum distance of 2 mm. Therefore, the maximum off-parallel difference between the two butt joint ends would be 2.5 mm and the minimum distance 2 mm over 50 mm and 5 mm and 4 mm respectively for a 100 mm length of butt joint.

16.6.3.3 Example Military Helicopter Common Control Unit

An example of a major manufacturing fault is for the common control unit, which is a key pad-screen interface into the military helicopter's computers. On operation of a single key press the screen went through two menu screens [key bounce] instead of only one. The fault was traced to a small pip (a small protrusion) on the underside of the key molding remaining from the manufacturing process. The corrective action was to remove the pip.

16.6.3.4 Example Military Helicopter Main Windscreen

Another was a military helicopter main windscreen that was made with six layers of alternate glass and polycarbonate. The screen is not a load carrying part of the airframe. One of the layers cracked in the hangar overnight. The fault was traced to the manufacturing process producing the screens on a horizontal jig. This process placed unwanted stresses into the screen when fitted to the airframe in the vertical position. The corrective action was to change the jig to the vertical position.

Also, large military products are usually integrated with the subsystem by the prime contractor. This introduces additional failure modes when the integration process fails and the system has failures when all the subsystem LRUs are functioning correctly but the system fails, i.e. the interface design documents are in error between the subsystems. This produces NFF by definition at the original equipment manufacturer (OEM).

16.6.3.5 Example Military Helicopter Tail Oscillation

Another examble was on a military helicopter that had an oscillation of the tail of the airframe during the development phase. The investigation looked into the stiffness of the airframe, the airflow around the airframe, hydraulic servos in the flight control system and the flight control auto-stabilization system. None of these were found to be at fault; they were all operating as designed and were built to the design interface specification from the system integrator. It was an integration fault. The corrective action, chosen by the system integrator, was to change the software in the flight control auto-stabilization system. This was also the cheapest solution.

16.6.3.6 Example Air-to-Air Missile

An example of a major integration fault was on the air-to-air missile. When testing the electronic units from different vendors together it was found that there were high electrical currents between the units. The fault was traced to the electronic units. One had been designed at 4.5 V whilst the other was designed for 5.0 V. Unfortunately, the design specification did not specify the power supply voltage.

16.7 Calculating in-Service Reliability

These examples show therefore that, these additional failure modes of extant design, manufacturing and subsystem integration faults have to be taken into account in the in-service reliability (R) measurement methods and the definition of NFF. Training analysis for reviewing component record cards and maintenance job cards for trending may be necessary.

Failure modes generated by use of the system outside its design envelope and any failure modes introduced by maintenance action together with any disturbance of a LRU/system, will degrade its reliability. These failure modes are not random and are strictly NFF and classed as non-attributable [4].

Failure modes introduced by maintenance action, that causes any disturbance of a LRU/system will degrade its reliability. The above are example of failure modes caused by design, manufacturing or integration issues, which will have corrective action devised and implemented. But, similar types of failure modes which are of a minor nature will be usually classed as ‘random’ and have no corrective actions devised. This suggests that a repairable system, e.g. a helicopter, will not have a constant hazard function. These failure modes are not random and are strictly NFF and should be classed as non-attributable.

When modelling the life-cycle costs (LCC) of a helicopter fleet it was common practice to ensured the reliability (R) used for LRU modelling included elements extant design/manufacture and non-attributable failure modes (i.e. design, manufacture or design). From experience it was observed that it was common practice for the helicopter mean time between failure (MTBF) to be reduced by up to 40% and there was a further reduction 15% to account for non-attributable failure modes. This 15% was considered to be the percentage generated by normal behavior of the helicopter operator.

The reason for this is, the support contract was fixed price for the cost of all repairs and new buy spares-scrapped items. Including these normal behavior, non-attributable failure modes ensures the risk to the prime contractor of losing money is minimized. As part of a support contract was in operation the normal behavior for non-attributable failure modes were monitored along with the design or manufacture failure modes to ensure compliance. Again, better trending of data may be required.

Table 16.1 Relationship between reliability and increased test period

	Inputs				Outputs		
	R	α	β	E	R_L	% NFF	Quantity NFF
Base [1]	0.8	0.05	0.05	0.8	0.948	21	31.5
Base [2]	0.8	0.05	0.05	0.9	0.968	19	28.5

The following examples demonstrate what happens if more time, money and effort were invested in the test equipment, manuals and engineers' training and the effect on R_L and NFF.

Table 16.1 shows the effect if the MT coverage (E) is increased from 0.8 to 0.9. It should be assumed that there are 150 maintenance tests in any period.

The results show an increase in reliability (R_L) from 0.948 to 0.968 and a decrease in NFF of 2%.

16.8 Conclusion

The MT model can be used when you need to assess whether to invest in better test equipment and engineers' training to reduce α and β probabilities. This would provide better test results and benefit the overall operational availability and reduce the occurrence of NFF. A good result for all involved.

Furthermore, the MT model can be used to analyze current maintenance scenario for their effectiveness.

NFF is related to E, α and R. Whereas R_L is related to E, β and R. Therefore, E has a major effect on the MT results and should therefore not be disregarded when considering system and component reliability.

Highly reliable systems may have a minimum time between MT to reduce the quantity of NFF. Due consideration should be given to the periodicity between each MT. Too frequent testing will increase the number of FAIL-FALSE and NFF. Consequently this will impact the operational availability.

References

1. Fedrick V (2014) Investigating the contribution of test and analysis to the no fault found (NFF) issue within the UK MOD. Thesis, Cranfield University
2. Hockley C, Lacey L, Pelham JG (2016) Report for air command on the impact of no-fault found on a selection of RAF aircraft fleets. EPSRC Report, Cranfield University
3. Jones JV (2006) Integrated logistic support handbook, 3rd edn. New York Sole Logistics Press, McGraw-Hill
4. Thompson J (1994) Mathematics of FRACAS, advances in logistics, Thesis. University of Exeter MIRCE research centre

Part VI

Cost, Obsolescence, Risk and TES Contract Design

Chapter 17
Best Practices in the Cost Engineering of Through-Life Engineering Services in Life Cycle Costing (LCC) and Design To Cost (DTC)

Paul Baguley

Abstract This chapter defines a number of Cost Engineering challenges from industry and their potential best practice solutions as industry case studies and industry practices surveys completed during the previous 5 years. In particular Life Cycle Costing in the context of upgrade and revamp in the process industry and also an example of design for full life cycle target cost for the manufacturing industry. Life Cycle Costing of complex long life cycle facilities is exemplified by identification and development of a life cycle costing of oil refineries through a survey of 15 companies and full life cycle experts and a review of the literature. Life cycle costing practices and a standardised life cycle cost breakdown structure are identified. Design to full life cycle target cost practices have been identified in the development of a full life cycle cost estimating tool for marine radar systems. In particular a survey of 17 companies and a case study with a marine radar systems company has identified specific practices useful in developing products to full life cycle target cost. In planning for future Through Life Engineering Services it is proposed that the collection of cost data and the understanding of Cost Engineering practices is a potential competitive advantage.

17.1 Introduction

Cost Engineering is a potential key enabler in Through Life Engineering Services. There are two systems considered in this research, one of Oil Refineries and one of Marine Radar Systems. The research provides examples of industry best practices from two industry surveys on best practices in the area of Design To Target Cost (DTC) and Life Cycle Costing (LCC). Each one contributes to the Through Life Engineering Services Function. Two tools have been developed to contribute to

---

P. Baguley (✉)
Centre for Innovative Manufacturing in Through-life Engineering Services,
Cranfield University, Cranfield, UK
e-mail: p.baguley@cranfield.ac.uk

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_17

decision making in this research. A Life Cycle Cost Breakdown (LCCB) Structure has been developed to specifically identify the life cycle cost concepts in an oil refinery relevant to Through Life Engineering Services. A Life Cycle Cost Estimation Software Tool Prototype was developed to provide cost information in decision making for design to Target full Life Cycle Cost for radar systems.

17.2 Contents

Section 3 states the aims and objectives of the research. Section 4 introduces life cycle costing concepts possible for Through Life Engineering Services. Section 4 describes an industry best practices survey in life cycle costing practices for oil refineries and a life cycle cost breakdown structure. Section 5 describes an industry best practices survey in manufacturing for Design to Target Full Life Cycle Cost. Sections 6 and 7 defines the development and structure of an industry proprietary Cost Estimation Decision Making tool to make design for target cost decisions. Section 8 discusses the potential benefits and drawbacks of the Cost Engineering concepts deployed in the Through Life Engineering Services context. Section 9 ends with conclusions.

17.3 Aim and Objectives of the Research

The aim of the chapter is to demonstrate the potential of Cost Engineering in Through Life Engineering Services, in particular Life Cycle Costing and Design to Target Life Cycle Costing.

The objectives of the chapter are to:

- Identify example best practices in life cycle costing for oil refineries in the literature and in industry
- Identify example best practices in design to Target Life Cycle Costing for manufacturing systems in the literature and in industry
- Define a prototype cost estimation tool for use in Design to Target Cost (DTC).

17.4 Life Cycle Costing in Oil Refineries Industrial Survey

A questionnaire was used to collect industry best practices about life cycle costing from a variety of existing industry contacts at Cranfield University. The purpose of the questionnaire is to collect information about the current life cycle costing practices in the oil refining industry. The information gathered assists in the development of a conceptual life cycle costing model and its cost breakdown structure as defined in Sect. 17.4.4.

17.4.1 Research Methodology and Questionnaire Design

The survey was carried out in five consecutive stages. Most of the questions were ‘open-ended’ as this gave the respondents the flexibility to respond without being restricted by the context of the questions.

Questions 1–7 were used to identify the role of the companies in the oil and gas industry, the kind of petroleum products they deliver, and the configurations and complexities of their units.

Questions 8–20 were used to gather information on current life cycle costing practice in the industry.

Questions 21–24 were used to solicit information on current operation and maintenance challenges being experienced in industry.

Questions 25–27 cover environmental impact issues and challenges.

Questions 28–30 were used to collect information on current risks and uncertainties associated with oil refinery life cycle costing.

It must be emphasised that the choice of respondents was done on purposive basis because of the high level of specialisation and professionalism required in this sector of the oil and gas industry [1]. The questionnaire was sent to a sample of 32 individuals and companies, all of which have been known to be experienced oil refiners, oil/chemical plant cost engineers, chemical engineers, design engineers, and independent consultants with interest in the life cycle costing of industrial plants. An introductory letter was written to solicit their participation in the study, as well as stating the overall goal of the survey. Following an information process check, 20 completed questionnaires were received but subsequent preliminary examination of the answers showed that usable responses found to be adequate for analysis amounted to 15. This corresponds to 47% of the total sample. This is unusually high because of the targeted nature of the survey. In line with the commitment given to the respondents, individuals and companies are not identified by name.

17.4.2 Life Cycle Costing in Oil Refineries Data Analysis

The results and analysis of every question will be presented in this section. It is important to state that not all respondents answered all the questions, and some answers are synonymous, hence are reworded to convey the same meaning. To increase the accuracy of the descriptive analysis of the results, the number of answers conveying the same meaning were categorised and put in parenthesis in a number of instances. The results of the survey have been arranged question by question.

Question 1: What sector of the oil and gas industry do your company operate?

Comments:

- Oil refining industry (7 respondents)
- Offshore/Upstream sector of the oil and gas industry (2 respondents)
- Industrial plants cost engineering (3 respondents)
- Design and project management (2 respondents)
- Power generation and chemical plant installation (1 respondent).

Question 2: What kind of products and services do you deliver?

Comments: Responses received include:

- Major petroleum products (6 respondents)
- Refinery decontamination chemicals and services (2 respondents)
- Consultancy related services (5 respondents)
- Offshore oil production and facilities maintenance (2 respondents).

Question 3: What are the main functions of your business?

Comments: This question is similar to Question 1, so almost all the respondents repeated the answers they gave in Question 1.

Question 4: What is the number of employees in your business unit?

Comments: Responses to this question varied according to the size of the organisations and consultancy outfits. The average number of employees for consultancy firms is 20 while the average number of employees in companies is 200.

Question 5: What is the average life expectancy of your plant?

Comments: Respondents from consultancy firms did not answer this question while almost all respondents involved in crude oil processing chose over 20 years as the lifespan of their plants or equipment. This means that the physical life of an oil refinery is above 20 years.

Question 6: What is the installed capacity of your refinery?

Comments: 80% of respondents involved in oil refining chose 100,000–150,000 bpd (barrels per day) as the installed capacity of their main unit. Refineries of this charge capacity are therefore common in UK and Nigeria.

Question 7: What is the level of complexity of your refinery?

Comments: Shows that 80% of the respondents involved in oil refining chose catalytic cracking refinery, 10% of the respondents chose hydro-skimming, 10% chose topping while no respondent opted for Coking refinery.

Question 8: What is your role in cost engineering in the oil and gas sector?

Comments: The percentage numbers of respondents and their roles are: equipment/spares procurement, maintenance costs and performance data reporting (47%); cost engineering consultancy services (33%); contract reviews and preparation of in-house estimates for new and existing facilities (13%); preparation of risk-based investment plans and models (7%).

Question 9: What do you consider to be the current challenges in oil refining and oil and gas industry?

Comments: The percentage numbers of responses according to the challenges are:

- Low capacity utilisation and rising cost of ownership (33%)
- Plant complexity and turnaround maintenance (20%)
- Non-availability of trained and experienced personnel to replace an aging work force (7%)
- Competition and dwindling profit (20%)
- Scope definition (7%)
- No response (13%).

Question 10: What do you understand to be Life Cycle Costing?

Comments: 80% of the respondents have basic knowledge of what life cycle costing means. The aggregation of their definitions implies that life cycle costing is the total cost of a product from conception to disposal.

Question 11: What methods do you use in life cycle costing?

Comments: 53% of the respondents acknowledged the existence of various investment appraisal methods that could be used by decision makers. The methods they presented ranged from net present value to cost benefit analysis. But net present value (NPV) is an economic evaluation method which is just a step among several steps to be undertaken in the life cycle costing analysis of a product, while cost benefit analysis is an evaluation method undertaken during the feasibility studies of new investments. The respondents from their answers do not have in-depth understanding of the life cycle costing methods. This implies that there is a lack of a standardised and normalised procedure that could be applied in the life cycle costing analysis of oil refineries. Hence, the standardisation of procedures is the main deficiency to be tackled.

Question 12: What data and information (sources) are used in life cycle costing?

Comments: 13% of the respondents answered this question. They said that the Cost Breakdown Structure, historical plant data, and corporate asset maintenance registers could be used as sources of data. The number of responses shows that the entrenchment of LCC techniques in the industry still appears to be insufficient.

Question 13: What are the challenges in life cycle costing?

Comments: 60% of the respondents answered this question. The answers are: poor asset historical data, uncertainty in performance, high cost of plant replacement, cost of revamping, and increased operation and maintenance cost. This implies that lack of historical plant data, uncertainties in plant performance (reliability and maintainability), cost of replacement or upgrading, revamping cost, and increased operation and maintenance costs are the challenges facing the industry and the successful implementation of life cycle costing.

Question 14: What is your understanding of the technological options in oil refining?

Comments: This question was completely misunderstood by the respondents. The author could have reframed it to convey its real meaning. However, the author meant “their understanding of refinery configurations”. I presume that Question 7 must have taken care of this question?

Question 15: Could you please describe the life cycle costing process? For instance, what are the steps? Do you have an example?

Comments: Most of the respondents repeated the answers they gave in Question 11. This question refers to the detailed steps to be undertaken in arriving at the life cycle cost of a product, which is more elaborate than just mentioning the conceptual life cycle costing model that shows cost categories in the life cycle costing process or framework. Notwithstanding the mix up, it was identified that no respondent made mention of a cost breakdown structure (CBS) which is the engine room of any life cycle costing analysis. The responses show that there is no standardised cost breakdown structure with the features needed for life cycle costing to be progressively executed. This implies that staff and departments responsible for evaluating investments in the oil refining industry lack a long-term perspective of asset management. The lack of a standardised CBS could make it impossible to conduct comparative analysis between different projects or to conduct single project analysis for budgetary purposes. A standardised CBS is therefore recommended for the industry.

Question 16: Please indicate the cost drivers you consider relevant for the life cycle costing of an oil refinery/oil and gas industrial assets?

Comments: 80% of the respondents answered this question. The responses include: plant investment; plant reliability and maintainability; plant complexity; energy; downtime; plant flexibility; and plant capacity. Hence, the aforementioned refinery cost contributors could be taken as the high level cost drivers.

Question 17: What are the relationships between the more significant ones?

Comments: 53% of the respondents said that reliability could drive down maintenance cost as presented. However, 47% of the respondents did not answer this question. Reliability as a matter of fact can reduce maintenance cost because if a plant is reliable the frequency of failure will be reduced thereby reducing maintenance cost.

Question 18: What are the life cycle stages of an oil refinery?

Comments: 80% of the respondents mentioned various life cycle stages with terminologies that could be categorised to portray the same meaning and provide standard life cycle stages for the oil refinery. For example, R&D, concept, and definition stages could be taken as a Research/Development Stage. Design/development, development, design, assessment, production, and manufacturing stages could be taken as Design/Manufacturing Stage. Investment, installation, acquisition, construction, and commissioning could be taken as Acquisition/Installation Stage. While in-service, facility usage, operation, maintenance, utilisation, and operation/support could be taken as Operation/Maintenance Stage. For the disposal stage, some respondents used retirement, end of life, recycle, remanufacture, and decommissioning. These stages could be categorised to mean Retirement/Disposal Stage.

Question 19: How many codes and standards of which the title includes the term “Life Cycle Costing” do you know?

Comments: Only 2 respondents answered this question. They mentioned PAS 55, ISO 15663, HM Treasury ‘Green Book’, and NATO/RTO Code of Practice for Life Cycle Costing. This means that most respondents are not aware of International Standards for Life Cycle Costing.

Question 20: How many of the codes and standards are specifically meant for the oil and gas industry?

Comments: There is no response to this question except one that mentioned ISO 15663—Petroleum and Natural Gas Industries: Life Cycle Costing Standard.

Question 21: What are the challenges in operation and maintenance?

Comments: 80% of the respondents gave the challenges as: lack of experienced staff, making value-based decisions on maintenance intervals, cost of maintenance, turnaround maintenance scheduling, and downtimes, while 20% of the respondents gave their challenges as technical and managerial problems. This implies that major operation and maintenance challenges in the industry are: expertise, mean-time-to-repair (Maintainability), reliability, routine maintenance planning, cost of lost production, and management policies.

Question 22: What are the issues in operation and maintenance related to life cycle cost?

Comments: This question is similar to the last question but with emphasis on life cycle costing. The answer given by 80% of the respondents includes: maintenance cost, spare parts availability, budget restrictions, increasing risk with declining condition, long lead items, downtimes (cost of lost production).

Question 23: What bottlenecks are there in operation and maintenance?

Comments: 60% of the respondents answered this question and gave the logjams as: resources, staff skills, and plant’s performance.

Question 24: What operations and maintenance models do you use? For example mathematical models, decision making models, or scheduling models.

Comments: 67% of the respondents answered this question. The responses include: Primavera planning/scheduling, and macro project models.

Question 25: What are the environmental impact challenges of CO_2 emission and its cost related issues?

Comments: 80% of the respondents answered this question. The responses centred on the topical issue of international legislation on the impact and cost of CO_2 emission (CO_2 taxes). From the responses it seems some companies are contemplating the inclusion of CO_2 cost into the design of new plants and cost models because of the international regulations on CO_2 emission.

Question 26: What are the technologies to curb environmental impact for now and in the future?

Comments: 60% of the respondents answered this question and gave the technologies as: carbon sequestration technology, flue gas desulphurisation. Carbon sequestration technology involves capturing CO_2 emitted from power plants and other industrial complexes and injecting it into geological structures deep below ground for long-term storage. The recovered CO_2 could be used for enhanced oil recovery (EOR) projects.

Question 27: What are the environmental impacts cost drivers and cost models?

Comments: 60% of the respondents mentioned environmental remediation cost while 40% gave CO_2 tax and health damages as cost drivers. However, they did not mention any cost model currently in use for the evaluation of environmental impacts.

Question 28: What are the significant risks associated with an oil refinery and appearing in the life cycle costing?

Comments: 47% of the respondents gave the associated risks as: plant upgrading and revamping, data availability, plant reliability, high investment cost while 53% of the respondents mentioned plant operation, maintenance, and environmental remediation cost as risks.

Question 29: What are the uncertainties in life cycle costing in refineries?

Comments: 47% of the respondents gave the uncertainties as plant lifespan, discount rates, energy cost while 40% mentioned data accuracy and estimating errors.

Question 30: What are the methods used to model risk and uncertainty?

Comments: Few companies and firms (20%) possess standardised procedure for evaluating risk analysis and uncertainty and this ranged from risk analysis based on individual task measurement, Monte Carlo simulation, and provision of a defined risk register. 60% lack a systematic procedure for this purpose. The responses to this question show that minimal use is made of risk and uncertainty estimation, and this could impinge on the full advantage that could be derived from the LCC technique.

17.4.3 Life Cycle Costing in Oil Refineries Questionnaire Summary

The results of the data analysis raised a vital issue about standardised procedures for the determination of a comprehensive life cycle cost analysis for oil refineries. The implications of the findings suggest that indeed there is a lack of a standard conceptual life cycle costing model with major cost categories and cost breakdown structure specifically designed for oil refineries. The recommendation is therefore for a standardised model and its cost breakdown structure to be integrated into an overall LCC framework.

17.4.4 Life Cycle Cost Breakdown Structure

Despite the existence of standards like PAS 55, there was not a useful reference Cost Breakdown Structure (CBS) for use in life cycle costing of oil refineries. From a literature review of 20 journals and the use of the industry survey in Sect. 17.4 then the Cost Breakdown Structure (CBS) was developed in Fig. 17.1. The CBS provides identification of significant costs for cost estimation, cost reduction or other potential Cost Engineering activities.

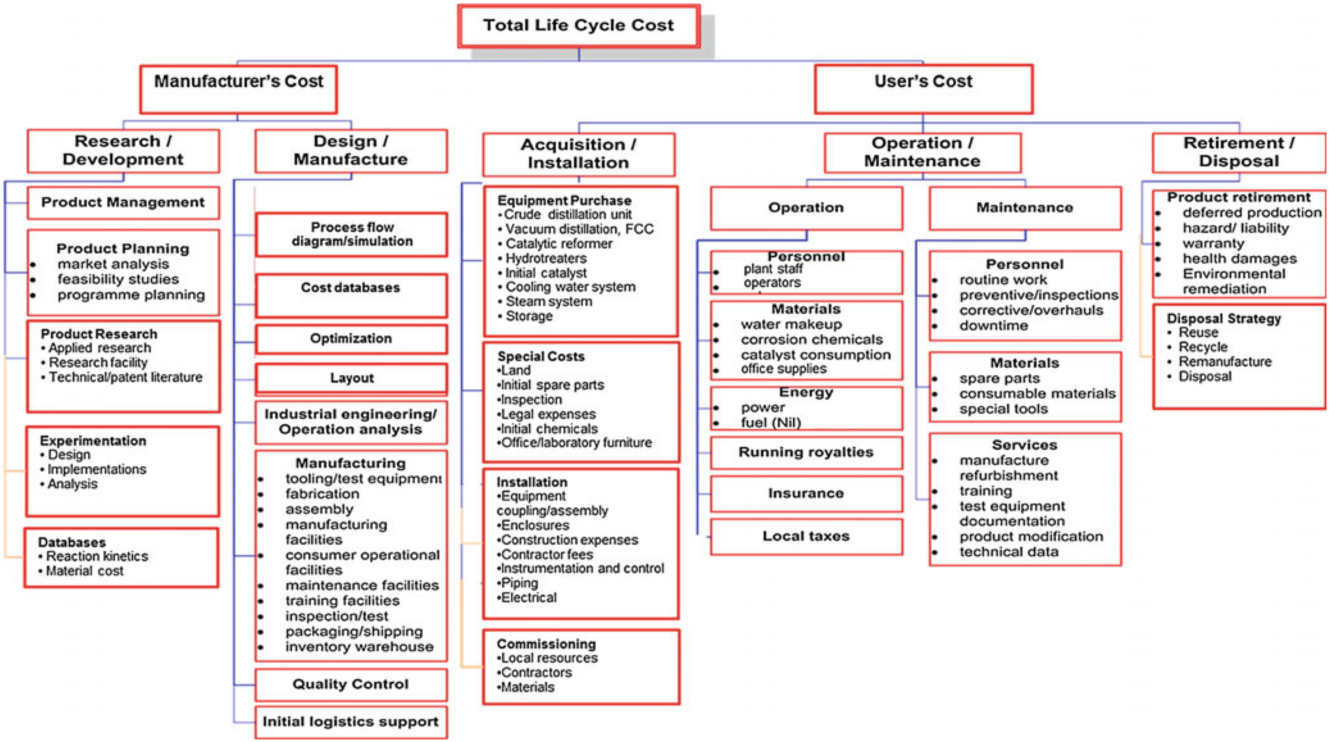


Fig. 17.1 Life cycle cost breakdown structure for oil refineries

17.5 Design to Full Life Cycle Target Cost Best Practices

The second part of this chapter is about design to target life cycle cost in a marine radar systems company. The research method was similar for the life cycle cost study in Sect. 17.4. However the individual best practices have been delineated, aggregated and summarised in the following section:

17.5.1 Results of the Design to Target Cost Survey

17.5.1.1 Design to Cost

Any product that is being designed, previously to the design phase it must be clear the features it will include, the value added of them and how it is possible to take into account all the costs. This is an important part because you need to know how you will deal with costs and if the level of detail requested can be achieved.

Then the Work Breakdown Structure (WBS) of the product must be detailed to its lowest level (e.g. to a component level for a software). DTC is not possible without a WBS.

Next step would be data collection. Lack of data was found to be a potentially common issue. In order to estimate or complete the data required for design to target cost task all the resources available must be used:

- Historical
- Experience
- Expert judgement
- Public domain data
- Supplier

Not always 100% data requested is available, so it can have a subjective part based on own experience or experts' opinion without losing validity.

The subsequent step is to define the appropriate Cost Estimation Methodology that will be used. If possible more than one methodology should be used and also cross data used to validate results. The most used in industry are detailed cost estimation, cost estimation by analogy and parametric cost estimation.

Parametric Cost Analysis potentially provides reasonable results for decision making. However sometimes it is too complex.

Analogy Cost Analysis is based on previous knowledge.

Detailed Cost Analysis is the easiest one, but is not always available, especially for a high level of detail.

It is important to model risk and uncertainty to improve the level of detail in the cost information. This cost information is identified by having a contingency cost. Fundamentally the process should begin with the identification of risks and then a process to find a way to mitigate these risks.

Best practice companies on DTC have been found to potentially use trade-off methodologies to assist in the decision making process during the design phase.

17.5.1.2 Life Cycle Cost

Best practice companies on Design To Target Cost use the Life Cycle Costing methodology. It is important to define the different Life Cycle Stages. This allows the knowledge of where major costs of the product are and then the subsequent decision making about cost. For some products it may be one stage that drives the major part of the Life Cycle Cost but even for those products it is needed to define the different stages to build a cost trade off-tool and determine what the optimal LCC is.

When all the stages are identified the Design Tom Target Cost process focusses on the one that drives the major part of the cost. The next stage is to define the cost drivers associated to these Life Cycle Stages. Build more accurate models for the most relevant stages.

17.5.1.3 Parametric Cost Estimation

Estimation techniques used in industry are:

- Analogy method. The analogy method is based on actual project data but cannot be applicable due to a new different technology.
- Detailed Cost Method. The detailed method may be laborious and time consuming, but it can result in a fairly accurate estimate if the work content is well understood.
- Parametric Method. The parametric estimation is a flexible and potentially reasonable methodology. However it can appear that it is a complex analysis not applicable to develop a new product.

Then in summary use a contingency cost that will cover variability in the estimation of costs, as all are estimations, the final cost will be different. Contingency cost should be a percentage of each of the costs estimated.

17.5.1.4 Cost Trade-Off

From the survey the best practices identified on the trade-off tools are:

- Choose the most cost effective technical solution.
- Identify all hidden costs such as quality. Quality should be treated as any other attribute, with an associated cost and that also adds value.
- Increase quality which leads to an increase in market share.
- Cost trade-off is many times used in the Make or Buy decision. It is suggested to keep competitive manufacturing in house.

17.6 Marine Radar Systems Full Life Cycle Costing System

A Cost Estimation tool was built for the Marine Radar System. Requirements were captured using face to face interviews with the senior managers and directors of the company, and prototyping during an agile development process.

The user interface was to allow an easy and user friendly interaction between the user and the tool. This graphical interface is developed in Microsoft Excel using VBA coding. To achieve that it is necessary to identify precisely who will be the end users of the tool. In this case it is quite broad as every employee of the company might use the tool. Therefore it is even more consequential that the interface is well informed with notes and accurate words and that it is aesthetically easy to understand.

Once the end user is identified, the global outputs expected by the user have to be defined. For the tool, the requirements were to calculate the new cost after making some changes on the product during the design phase. This is done via use cases.

To make sure the requirements are met, the user interface can be updated and improved through a precisely commented code. This code is using data from the database, as well as user inputs, to make calculations depending on the changes made from the former configuration of the product in order to get the costs of the new product.

17.7 Parametric Design to Target Cost Components and Architecture

The requirements led to the following design of the Parametric Cost Model Development Tool. The software tool system includes inputs, processes, outputs and system constraints for the proposed system. The main components can be considered to be:

Inputs as selections

- Type of product, starting product(s), ways of connecting multiple products; building blocks and components of a product present in the new configuration, i.e. a Bill of Materials; building blocks and components that will be studied during the cost estimate.
- Level of the cost estimation (building block level or component level); the estimation method that will be used for each unit; inputs as values; percentage of similarity at the “analogic estimation” method; cost driver value at the “parametric estimation” method; actual cost at the “detailed cost” method.

Processes

- “Analogic estimation” method; “Parametric estimation” method; “Detailed cost” method; adding new products, building blocks and components at the database; modifying existing units in the database.

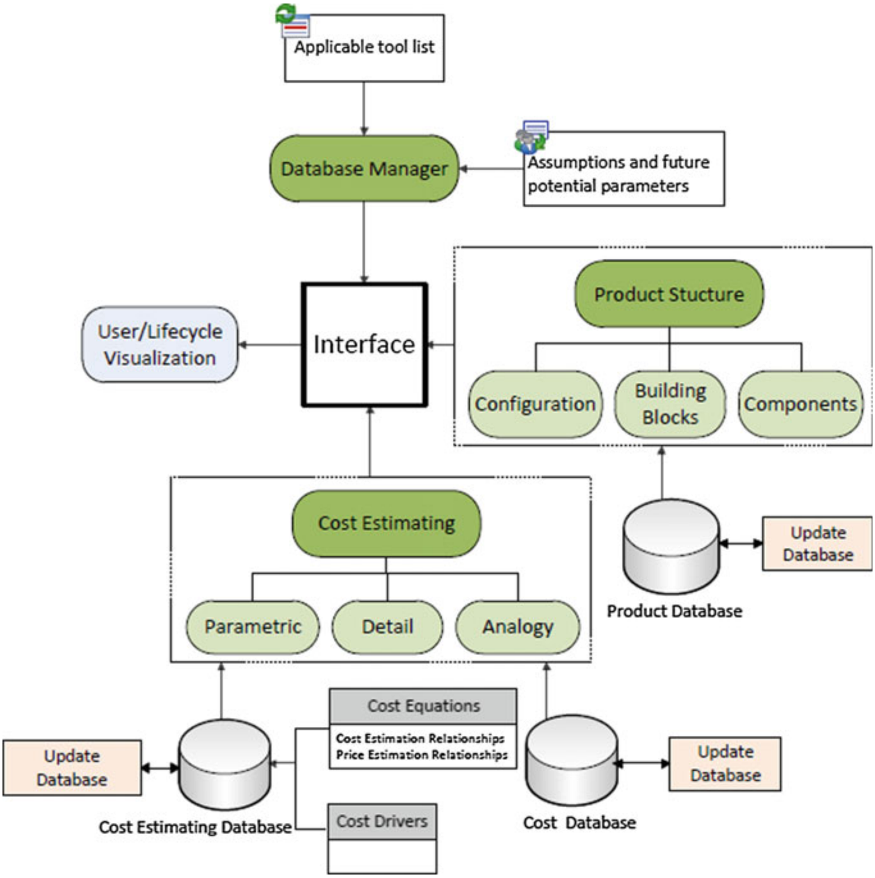


Fig. 17.2 Diagram of the design to target cost tool

Outputs

- Total cost of the starting and the new product configuration; Cost of each lifecycle stage for the starting and the new product configuration; contribution percentage of each lifecycle stage to the total cost.

System constraints

- The tool needs to use an existing product configuration as a starting point; specific fields in the database must be complete in order to use any of the estimating methods; system maintenance will be needed in order to ensure the continuous effectiveness of the tool and the accuracy its parametric provide.

The architecture of the final tool can be seen in Fig. 17.2. Fundamentally three cost estimation philosophies are deployed using a cost estimating database. The parametric cost equations are developed and stored with the data in the database.

17.8 Discussion

Life Cycle Costing is a process which has potential for improvement by scientific methods in the oil refinery business as shown by the simple and limited methods currently in operation. There is a requirement to improve cost reduction for the operating costs to improve the case for asset management. Understanding of cost drivers will aid in the calculation of availability. Reliability is of main importance in oil refineries, in particular loss of availability is of more importance than cost. Design to Target Cost is a novel Cost Engineering philosophy in the Marine Radar Systems industry. In the case study in particular there was no specific Cost Engineering capability in the company. The project identified a software based cost estimating system, utilising all three philosophies of detailed cost estimation, cost estimation by analogy and parametric cost estimation. The centralised collection of cost data was a main novelty and provided a significant improvement to decision making. Indeed the main problem in the company was cost effective decision making in new product development in the context of through life engineering services.

In the respect of the Design To Cost (DTC) concept then it was found in the industrial case study that these ideas were of importance in generating the design to cost intervention. Namely how to organise for Design To Cost (DTC), the cost of Design To Cost, what cost estimation tools support Design To Cost (DTC), what are the Design To Cost (DTC) activities in the process of Design To Cost (DTC), who are the stakeholders involved in the Design To Cost (DTC) process.

Organising for DTC involved the concepts of an organisational hierarchy occurring at customer level, director level and manager level. Stakeholders occurring below this hierarchy level are Engineering, Manufacturing, Procurement, Logistics, Finance and Procurement for instance.

There are fundamental philosophies of cost estimation, in particular detailed cost estimation, costing by analogy, parametric cost estimation and expert judgement. Design to Cost and Design to Target Cost require cost information during the full life cycle of the product including the use and disposal stages. These latter stages have historically not facilitated the collection of cost information in order to make cost predictions. Therefore cost estimation philosophies like parametric cost modelling or expert judgement are the only tools which can be potentially effective in this space. In the case study in this research it was decided to plan to develop all cost estimation philosophies into one tool as a proprietary solution.

Cost effectiveness in Design To Cost (DTC) is found in several areas. The capability to collect cost information and make predictions during the Design To Target Cost process is the significant cost driver in cost estimation. Accuracy is proportional to the amount of information available in order to make cost predictions.

Cost information collection from suppliers introduces the problem space of cost management. It is actually price information that is collected. In this case study there were minimal data points available in the form of price data as the homogeneous product data sets were of limited size and range. This meant that in the prototype proprietary system cost models were only able to provide indicative trends and in a visual way. However a start had to be made in order to build future systems.

The Design To Cost (DTC) activities which were of main importance in the case study were the independent use of the proprietary cost estimation tool by stakeholders. Previously cost related decisions were expensive since they involved senior level decision makers and had to be appropriately coordinated. In addition lack of a cost estimation process, cost information and use of subjective judgement are the main points to be addressed and solved in the current design to target cost process.

The general concepts discussed in design to cost are about methodology, roles, budget, risk management, data collection and cost breakdown structure.

17.9 Conclusions

The advent of Through Life Engineering Services has meant a new and novel requirement for Cost Engineering research. This is because the service presents novel scenarios with novel cost information required. The example marine radar sector although possesses supplier data for manufacturing cost still requires expert judgement about full life cycle cost like obsolescence cost. The larger complex oil refinery industry lacked cost engineering tools from a more scientific background but understood that reliability and availability were significant drivers over cost. Best practices in Cost Engineering from industry show a fundamentally basic level of capability currently being used in recent Through Life Engineering Services. The indication is of a low level of maturity and indicates a potential opportunity to improve new services. Data collection and database development is a part of capability which introduces problems in uncertainties and risks in cost estimation and provides the problem of cost effectiveness in developing cost estimates for robust decision making in TLES. Because TLES can be a new concept for companies then design to target cost industry best practices and a database and proprietary cost estimation tool have been found to be useful capability for initial improvement. However it is not clear at the moment what a long term roadmap might be. That is in the context of a company which competes on full life cycle cost elements but does not necessarily contract for them. In large complex industries cost elements are significant and subject to significant variation. Life cycle cost is an important consideration for upgrade and revamp; however reliability is known to be the critical factor.

Reference

1. Robson C (2002) Real world research: a resource for social scientists and practitioners-researchers, 2nd edn. Blackwell Publishers Ltd., UK

Chapter 18
Cost Model for Assessing Losses to Avionics Suppliers During Warranty Period

Ahmed Raza and Vladimir Ulansky

Abstract Reduction of warranty maintenance costs is a critical issue to the manufacturers of avionic products. A method to reduce expected warranty costs is the determination of all components of financial losses to avionic product suppliers during the warranty period with further minimisation of these losses. This study interlinks the warranty, reliability and maintenance indicators of avionic products. Mathematical models are proposed for analysing and assessing financial costs to avionic system suppliers during the warranty period. The developed mathematical models consider the warranty period, reliability indicators with respect to permanent and intermittent failures, redundancy, number of spare parts, cost of restoration and transportation and penalties for exceeding the duration of warranty repair or replacement. Numerical examples illustrating the proposed models are provided.

18.1 Introduction

Efficient operation of an aircraft is largely determined by the proper regulation of the relationship between an aircraft supplier (manufacturer) and aircraft operator (airline). The most acute problem faced in this relationship is during the warranty period, where the main cost for aircraft systems' repair is borne by the supplier, i.e. the supplier will remedy the defect in a system free of charge and in a reasonable time, by either repairing or replacing the defective system. In the process of meeting

---

A. Raza
Overseas Projects and Maintenance, Department of the President's Affairs,
P.O. Box 372, Abu Dhabi, UAE
e-mail: ahmed_awan@hotmail.com

V. Ulansky (✉)
National Aviation University, Kosmonavta Komarova Avenue 1,
Kiev 03058, Ukraine
e-mail: vulanskyi@yahoo.com

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_18

the terms of a contract for the supply of aircrafts, it is necessary to choose a warranty maintenance strategy and to evaluate all the penalties that may be presented to the supplier in the event of non-compliance with the warranty. The regularity of aircraft flights and the costs that may be incurred by both sides will depend largely on the chosen strategy of warranty maintenance. Thus, the most important task that must be performed during the process of purchasing and commissioning new aircraft is the justification of the warranty maintenance and repair strategy and assessment of the supplier costs during the warranty period. This problem is topical because, for example, in 2015, aerospace product warranty claims paid by U.S.-based companies amounted to $1.6 billion [1]. The warranty policies are classified as free replacement warranty (FRW) and pro-rata warranty (PRW). Under the FRW, the manufacturer takes the responsibility to repair (or replace) any failed product free of charge during the warranty period. The PRW policy implies that the buyer shares the repair cost with the manufacturer during the warranty period. In this study, we consider only FRW policies. Modelling of FRW policies has received a lot of attention in the literature. The Product Warranty Handbook edited by Blischke and Murthy [2] is a collection of research papers devoted to warranty cost models. A review of the literature on warranty models and analysis methods was conducted by Thomas and Rao [3], wherein they analysed different FRW cost models. Chukova and Hayakawa [4] employed an alternating renewal process to model the operating and repair times and evaluated the warranty costs over the warranty period for FRW. Jun and Hoang [5] considered the discounted warranty cost model for repairable series systems under FRW policy. The proposed approach incorporates the system structure information, the value of time and the impact of repair actions. Huang et al. [6] evaluated the problem of estimating the expected warranty costs in case of intermittent and heterogeneous usage intensity. FRW policies for repairable and non-repairable products were analysed. Wang [7] evaluated four new warranty cost models including imperfect repairs. Warranty of a k-out-of-n system with imperfect maintenance was also analysed. Park [8] studied warranty cost models based on the quasi-renewal processes, exponential distribution and assumption that a repair service is imperfect. Blischke et al. [9] presented an integrated compilation of existing literature on warranty data collection. Cost models for one- and two-dimensional warranties were considered. The most important warranty cost models were analysed by Diaz et al. [10]. A method for minimizing additional costs of warranty service through optimal service strategies and efficient management logistics was shown. Shafiee and Chukova [11] conducted a systematic review of mathematical models, where the warranty is interlinked with the maintenance strategy. The research directions in warranty and maintenance were outlined, and a list of new and challenging topics were identified.

It should be noted that the considered models do not take into account the features of avionic systems' operation and maintenance.

18.2 Objectives of the Study

Avionics represent a significant component of aircraft costs: up to 40% in civil aircraft and more than 50% in military aircraft [12]. The complexity in avionics is continuously growing. Most avionic systems have a block-modular structure. Each of the systems of the flight control and navigation complex is a redundant system consisting of two or three line replaceable units (LRUs). In turn, the LRU includes a set of easily removable shop replaceable units (SRU), which are typical replacement elements for the LRU. Each LRU has a built-in test equipment (BITE), which monitors the operability of the LRU. In accordance with the architecture of avionic systems, there are three possible levels of maintenance. The upper level is the field or flight line maintenance, where the LRU can be removed and replaced if it was rejected by the BITE during flight. At the intermediate-level (I-Level) of maintenance, the failed SRUs are replaced to restore the operability of dismantled LRUs. The third level of maintenance consists of the restoration of SRUs by replacement of the failed chips at the depot or by the supplier. In some cases, to reduce the cost of maintenance, the intermediate level is eliminated [13]. This means that LRUs removed on the flight line are sent directly to the depot or supplier instead of heading first to the intermediate level. In this case, the major element of cost savings comes from the reduction of maintenance personnel at the intermediate (base) level [14]. Different strategies of warranty maintenance may differ primarily in the warehouse spare part system, presence of automatic test equipment (ATE) and the depth of the recovery of failed systems at the buyer's end. Obviously, the mathematical models that describe the total supplier cost should take into account the chosen warranty maintenance strategy. A high level of flight safety is ensured by the redundancy of avionic systems. Therefore, when evaluating the operating costs due to warranty claims, we should take into account the type of redundancy for each avionic system. Thus, the purpose of this study is to develop mathematical models for evaluating the financial costs to avionic system suppliers during the warranty period, taking into account the above-mentioned features of avionics architecture, maintenance and operation.

18.3 Analysis of Costs Components of Avionics Suppliers

An analysis of the warranty obligations of the avionics suppliers allows us to identify the following components of the financial costs: cost of repair or replacement of failed LRUs; penalties for delays in repairing or replacement of failed LRUs; capital expenditures related to the chosen variant of the warranty maintenance and operating costs; costs associated with the buy-back of the excess spare parts from the airline at the end of the warranty period and transportation costs.

The warranty period T_W can be represented in flight hours T or in calendar duration T_0 (years). In some cases, the length of the warranty period is specified separately in calendar duration and in the number of flight hours. In such cases, the warranty period is equal to the value, whichever is reached first.

Suppliers of avionics products guarantee the restoration or replacement of the failed items at their own expense. The repair of the defective product can be carried out by the supplier, as well as at certified repair centres. In any case, the supplier pays for the restoration. The penalty for any delay in delivery of repaired or replaced avionics product is associated with aircraft downtime due to the failure of a product under warranty from supplier. The amount of the penalty for each day (hour) of downtime can either be specified in the contract between the supplier and the buyer, or can be equal to the rent paid towards borrowing the missing items from other airlines. Capital expenditure related to the chosen variant of warranty maintenance and operating costs are associated with the presence of the supplier representatives at the buyer, as well as with the possibility of repairing the defective product at the buyer end. The possibility of buying the excessive spare parts should be stipulated in the contract between the supplier and the buyer. To reduce this loss component, the supplier should use accurate data on the reliability characteristics of the avionics products and operating conditions while calculating the optimal number of spare parts. Transportation costs of the supplier are primarily associated with the delivery of repaired items and spare parts to the buyer and shipping the failed items to the repair facilities. An analysis of the sales contracts for avionics products shows that the transportation costs can be shared between the buyer and the supplier. The contract usually specifies the party responsible for bearing transportation costs.

Thus, the financial costs to the supplier during the warranty period are defined by the following formula:

$$C_S(T_W) = R(T_W) + P(T_W) + C_E(T_W) + C_{SP}(T_W) + C_{TC}(T_W) \tag{18.1}$$

where $R(T_W)$ is the cost of repair or replacement of the failed LRUs during the warranty period, $P(T_W)$ is the penalty for exceeding the duration of the warranty repair or replacement within the warranty period, $C_E(T_W)$ is the capital expenditure related to the chosen variant of warranty maintenance and corresponding operating costs, $C_{SP}(T_W)$ is the cost of purchasing the excess spare parts for the LRUs from the buyer at the end of the warranty period and $C_{TC}(T_W)$ is the total transport cost borne by the supplier during the warranty period.

To identify the components of supplier losses we need to model the process of operation and maintenance of avionic systems.

18.4 Maintenance Model During the Warranty Period

The avionics LRU failures can be classified into permanent and intermittent failures. The condition of each LRU is continuously tested by its BITE and in case of a permanent failure, the LRU is switched off. If an intermittent failure occurs during flight, the LRU is usually not switched off but the on-board computer records the information on such events. The following procedure is set for recovery operations during the warranty period. If a permanent or intermittent failure occurred during a flight, the LRU is dismantled after landing and directed to the supplier for repair. We assume that the LRU becomes as good as new after repair.

18.4.1 Probabilities of LRU Repair

Consider an LRU that should operate for a finite time interval T, which is the warranty period. Assume that a random variable Ξ ($\Xi \geq 0$) denotes the time to a permanent failure, with a failure density function $\omega(\xi)$. Let us introduce the following notations for possible restoration of the LRU at time $j\tau$: $A(j\tau)$ is the event consisting of restoration of the LRU at time $j\tau$ after the j-th flight; $H_{IF}(j\tau)$ and $H_{PF}(j\tau)$ are the events corresponding to the restoration of the LRU after intermittent and permanent failure, respectively; $P_R(j\tau)$, $P_{IF}(j\tau)$ and $P_{PF}(j\tau)$ are, respectively, the probabilities of the events $A(j\tau)$, $H_{IF}(j\tau)$ and $H_{PF}(j\tau)$, where τ is the mean time between aircraft landings at the base airport. To determine $P_R(j\tau)$, $P_{IF}(j\tau)$ and $P_{PF}(j\tau)$, we introduce the probability distribution function (PDF) of random variables Ξ, Θ_1, …, Θ_j, which we denote as $\Omega(\xi, \theta_1, \ldots, \theta_j)$, where Θ is the time to intermittent failure with PDF $\psi(\theta)$ and

$$\Theta_j = \Theta - (j - 1)\tau \tag{18.2}$$

is the remainder of the operating time to intermittent failure after j–1 flights (j = 1, 2, …). Using the multiplication theorem of PDFs, we get [15]

$$\Omega\left(\xi, \theta_1, \ldots, \theta_j\right) = \omega(\xi)\Omega_0\left(\theta_1, \ldots, \theta_j|\xi\right) \tag{18.3}$$

where, $\Omega_0(\theta_1, \ldots, \theta_j|\xi)$ is the conditional PDF of random variables Θ_1, …, Θ_j under the condition that $\Xi = \xi$.

Further, we use two conditional probabilities associated with intermittent failures. The conditional probability of an intermittent failure occurring during the j-th (j = 1, 2, …) flight, under the condition that $\Xi = \xi > j\tau$, is formulated as follows:

$$P_{IF|\xi}\left[\overline{\tau, (j-1)\tau};j\tau|\xi\right] = P\left[\bigcap_{i=1}^{j-1}\Theta_i > \tau \bigcap \Theta_j < \tau|\xi\right] \tag{18.4}$$

The conditional probability of an intermittent failure not occurring during the j-th flight is stated as follows:

$$P_{\overline{IF}|\xi}\left[\overline{\tau,(j-1)\tau};j\tau|\xi\right]=P\left(\bigcap_{i=1}^{j}\Theta_i>\tau|\xi\right)\tag{18.5}$$

Since a random variable Θ_i (i = 1, …, j) is defined in the finite time interval (0, $T - (i - 1)\tau$], we need to introduce the conditional PDF $\Omega_1\{\theta_1, \ldots, \theta_j \mid \xi\cap[0 < \Theta_i \leq T - (i-1)\tau,\ i = 1, \ldots, j]\}$, which is expressed through the conditional PDF $\Omega_0(\theta_1, \ldots, \theta_j|\xi)$ as follows:

$$\Omega_1\left\{\theta_1,\ldots,\theta_j|\xi\bigcap\left[0<\Theta_i\leq T-(i-1)\tau,i=\overline{1,j}\right]\right\}=$$
$$\Omega_0(\theta_1,\ldots,\theta_j|\xi)\Big/\int\limits_0^T\int\limits_0^{T-\tau}\cdots\int\limits_0^{T-(j-1)\tau}\Omega_0(u_1,\ldots,u_j\,|\xi)du_1\ldots du_j\tag{18.6}$$

By integrating (18.6) over the corresponding limits, we determine the probabilities (18.4) and (18.5):

$$P_{IF|\xi}\left[\overline{\tau,(j-1)\tau};j\tau|\xi\right]=\frac{\int\limits_\tau^T\int\limits_\tau^{T-\tau}\cdots\int\limits_\tau^{T-(j-2)\tau}\int\limits_0^\tau\Omega_0(u_1,\ldots,u_j|\xi)du_1\ldots du_j}{\int\limits_0^T\int\limits_0^{T-\tau}\cdots\int\limits_0^{T-(j-1)\tau}\Omega_0(u_1,\ldots,u_j|\xi)du_1\ldots du_j}\tag{18.7}$$

$$P_{\overline{IF}|\xi}\left[\overline{\tau,(j-1)\tau};j\tau|\xi\right]=\frac{\int\limits_\tau^T\int\limits_\tau^{T-\tau}\cdots\int\limits_\tau^{T-(j-1)\tau}\Omega_0(u_1,\ldots,u_j|\xi)du_1\ldots du_j}{\int\limits_0^T\int\limits_0^{T-\tau}\cdots\int\limits_0^{T-(j-1)\tau}\Omega_0(u_1,\ldots,u_j|\xi)du_1\ldots du_j}\tag{18.8}$$

The probabilities of the LRU restoration due to the occurrence of intermittent or permanent failure are formulated as follows:

$$P_{IF}(j\tau)=P[H_{IF}(j\tau)]=P\left\{\bigcup_{\nu=0}^{j-1}\left\{A(\nu\tau)\bigcap\{B_1[(j-\nu)\tau]\backslash B_2[(j-\nu)\tau]\}\right\}\right\}\tag{18.9}$$

$$P_{PF}(j\tau)=P[H_{PF}(j\tau)]=1-P\left\{\bigcup_{\nu=0}^{j-1}\left\{A(\nu\tau)\bigcap B_1[(j-\nu)\tau]\right\}\right\}\tag{18.10}$$

where

$$B_1[(j-\nu)\tau] = [\Xi > (j-\nu)\tau] \bigcap \left(\bigcap_{i=\nu+1}^{j-1} \Theta_i > \tau \right) \tag{18.11}$$

is the joint occurrence of the following events: the LRU begins to work at time t_ν; it does not fail up to time $j\tau$ and no intermittent failure occurs during the flights $\nu + 1, \ldots, j - 1$;

$$B_2[(j-\nu)\tau] = [\Xi > (j-\nu)\tau] \bigcap \left(\bigcap_{i=\nu+1}^{j} \Theta_i > \tau \right) \tag{18.12}$$

is the event different from $B_1[(j-\nu)\tau]$ only in the fact that in the j-th flight also there was no intermittent failure;

\ is the symbol denoting the difference between the two events.

The LRU will be restored at time $j\tau$ if either of the events $H_{IF}(j\tau)$ or $H_{PF}(j\tau)$ occurs. Therefore,

$$A(j\tau) = H_{IF}(j\tau) + H_{PF}(j\tau) \tag{18.13}$$

Assuming that $H_{IF}(j\tau)$ and $H_{PF}(j\tau)$ are mutually exclusive events and applying this condition to (18.13), the addition theorem of probability gives

$$P_R(j\tau) = P[A(j\tau)] = P_{IF}(j\tau) + P_{PF}(j\tau) \tag{18.14}$$

If the LRU was restored at time $\nu\tau$, random variable Ξ is defined in the interval $(0, T - \nu\tau]$ with conditional PDF

$$\omega(\xi | 0 < \Xi \le T - \nu\tau) = \omega(\xi) \Big/ \int_0^{T-\nu\tau} \omega(x)dx \tag{18.15}$$

Taking into account (18.15) and applying to (18.9)–(18.12), the addition and multiplication theorems of probability, we obtain

$$P_{IF}(j\tau) = \sum_{\nu=0}^{j-1} \frac{P_R(\nu\tau)}{\int_0^{T-\nu\tau} \omega(x)dx} \int_{(j-\nu)\tau}^{T} P_{IF|\xi}\left(\overline{\tau, (j-\nu-1)\tau}; (j-\nu)\tau | \vartheta\right) \omega(\vartheta) d\vartheta \tag{18.16}$$

$$P_{PF}(j\tau) = 1 - \sum_{\nu=0}^{j-1} \frac{P_R(\nu\tau)}{\int\limits_0^{T-\nu\tau} \omega(x)dx} \int\limits_{(j-\nu)\tau}^{T} P_{\overline{IF}|\xi}\left(\overline{\tau, (j-\nu-2)\tau}; (j-\nu-1)\tau|\vartheta\right)\omega(\vartheta)d\vartheta \tag{18.17}$$

where $P_R(0) = 1$, $P_{IF}(0) = 0$ and $P_{PF}(0) = 1$.

As is well known, the exponential distribution provides an appropriate distribution of permanent failures of complex systems [16]. It has been reported in several publications that the exponential distribution is also an appropriate distribution for avionics products [17, 18]. This is because LRUs in modern avionics consist of a large number of electronic chips. A flight director system may consist of 460 digital integrated circuits (ICs), 97 linear ICs, 34 memories, 25 ASICs and 7 processors [18]. For these components, external failure mechanisms (electrical overstress, electrostatic discharge and so on) and intrinsic failure mechanisms (dielectric breakdown, electromigration and hot carrier injection) can cause the components to fail. Different failure modes contribute to a constant LRU failure rate. This is possible only with the exponential distribution of failure over time

$$\omega(t) = \lambda e^{-\lambda t} \tag{18.18}$$

where λ is the permanent failure rate of LRU.

Assume that intermittent failures are also subject to the exponential law with PDF

$$\psi(t) = \theta e^{-\theta t} \tag{18.19}$$

As is well known, the exponential distribution has the memoryless property. Therefore, conditional probabilities (18.7) and (18.8) are converted to the following form:

$$P_{IF|\xi}\left[\overline{\tau, (j-1)\tau}; j\tau|\xi\right] = \left(1 - e^{-\theta\tau}\right)\prod_{i=1}^{j-1}\left\{e^{-\theta\tau} - e^{-\theta[T-(i-1)\tau]}\right\} \Big/ \prod_{i=1}^{j}\left\{1 - e^{-\theta[T-(i-1)\tau]}\right\} \tag{18.20}$$

$$P_{\overline{IF}|\xi}\left[\overline{\tau, (j-1)\tau}; j\tau|\xi\right] = \prod_{i=1}^{j}\left\{e^{-\theta\tau} - e^{-\theta[T-(i-1)\tau]}\right\} \Big/ \prod_{i=1}^{j}\left\{1 - e^{-\theta[T-(i-1)\tau]}\right\} \tag{18.21}$$

Substituting (18.18), (18.20) and (18.21) into (18.16) and (18.17) we obtain

$$P_{IF}(j\tau) = \left(1 - e^{-\theta\tau}\right) \sum_{\nu=0}^{j-1} P_R(\nu\tau) \frac{\left[e^{-\lambda(j-\nu)\tau} - e^{-\lambda T}\right] \prod\limits_{i=\nu+1}^{j-1} \left\{e^{-\theta\tau} - e^{-\theta[T-(i-1)\tau]}\right\}}{\left[1 - e^{-\lambda(T-\nu\tau)}\right] \prod\limits_{i=\nu+1}^{j} \left\{1 - e^{-\theta[T-(i-1)\tau]}\right\}} \tag{18.22}$$

$$P_{PF}(j\tau) = 1 - \sum_{\nu=0}^{j-1} P_R(\nu\tau) \frac{\left[e^{-\lambda(j-\nu)\tau} - e^{-\lambda T}\right] \prod\limits_{i=\nu+1}^{j-1} \left\{e^{-\theta\tau} - e^{-\theta[T-(i-1)\tau]}\right\}}{\left[1 - e^{-\lambda(T-\nu\tau)}\right] \prod\limits_{i=\nu+1}^{j-1} \left\{1 - e^{-\theta[T-(i-1)\tau]}\right\}} \tag{18.23}$$

It should be noted that beginning from the fourth or fifth flight probabilities (18.22) and (18.23) usually reach the steady-state values

$$P_{IF}^{*}(\tau) = \left(1 - e^{-\theta\tau}\right)\left(e^{-\lambda\tau} - e^{-\lambda T}\right) / \left[\left(1 - e^{-\lambda T}\right)\left(1 - e^{-\theta T}\right)\right] \tag{18.24}$$

$$P_{PF}^{*}(\tau) = 1 - \left(e^{-\lambda\tau} - e^{-\lambda T}\right) / \left(1 - e^{-\lambda T}\right) \tag{18.25}$$

18.4.2 Expected Repair Costs

Expected repair costs during the warranty period are determined as follows:

$$R(T) = mN\left\{ C_{IF} \sum_{j=1}^{[T/\tau]} P_{IF}(j\tau) + C_{PF} \sum_{j=1}^{[T/\tau]} P_{PF}(j\tau) \right\} \tag{18.26}$$

where m is the number of identical LRUs in a redundant avionic system, N is the number of aircraft under warranty of supplier, C_{IF} and C_{PF} are, respectively, the mean cost of repairing LRU with intermittent and permanent failures by the supplier and $[T/\tau]$ is the integer number of ratio T/τ.

Equation (18.26) is simplified to

$$R(T) = \frac{mNT}{\tau}\left[C_{IF}P_{IF}^{*}(\tau) + C_{PF}P_{PF}^{*}(\tau)\right] \tag{18.27}$$

if the steady-state values of probabilities $P_{IF}(j\tau)$ and $P_{PF}(j\tau)$ are used.

Example 1 In modern wide-body aircraft there are usually three instrument landing systems (ILS), i.e. $m = 3$. Assume that $N = 4$, $T = 10{,}000$ flight hours, $\tau = 5$ h, $\lambda = 10^{-4}\ \text{h}^{-1}$, $C_{IF} = £1{,}000$ and $C_{PF} = £2{,}000$. Using (18.24) and (18.27), $P_{IF}^{*}(\tau)$ and $R(T)$ are calculated as a function of θ.

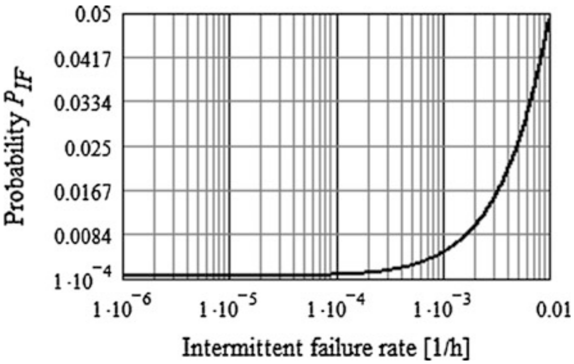


Fig. 18.1 Dependence of probability $P_{IF}^{*}(\tau)$ on intermittent failure rate

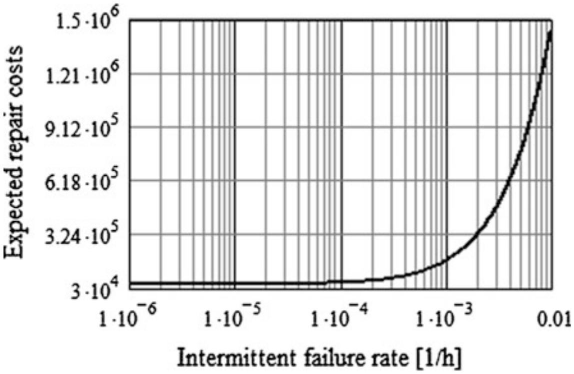


Fig. 18.2 Dependence of expected repair costs on intermittent failure rate

The dependency of $P_{IF}^{*}(\tau)$ and $R(T)$ as a function of θ are shown in Figs. 18.1 and 18.2, respectively. As can be seen from Fig. 18.1, the probability $P_{IF}^{*}(\tau)$ is 7.9×10^{-4} when $\theta < 9 \times 10^{-5}$ h^{-1}, and it begins to rise sharply when $\theta > 10^{-4}$ h^{-1}. From Fig. 18.2, it follows that the expected repair cost is £60,700 when $\theta < 8 \times 10^{-5}$ h^{-1}. However, when $\theta > 10^{-4}$ h^{-1}, the shape of $R(T)$ is similar to the shape of $P_{IF}^{*}(\tau)$.

Thus, the expected costs of warranty repairs may greatly depend on the intermittent failure rate.

18.4.3 Expected Penalty Costs

Optimal number of spare LRUs largely depends on the warranty repair time of defective LRUs. When the duration of the warranty repair time is exceeded, one of the following situations may take place.

Situation 1. There is at least one spare LRU in the warehouse. All aircraft under warranty by the supplier fly on schedule.

Situation 2. There are no spare LRUs in the warehouse. All aircraft under warranty of the supplier fly on schedule.
Situation 3. There are no spare LRUs in the warehouse. At least one aircraft is standing idle on the ground, while awaiting emergency delivery of spare LRUs.

In the first or second situation, the aircraft owner does not bear financial losses. When the third situation arises, the aircraft owner incurs losses due to violation of the schedule, flight cancellation, compensation to passengers, etc. If the supplier makes an expedited delivery under existing contract t_{ed}, it does not lead to financial losses. However, if the spare LRU is delivered after a time $t_{ed} + \Delta t_{delay}$, the supplier bears financial losses for each unit of time interval (t_{ed}, $t_{ed} + \Delta t_{delay}$). These financial losses may be equal either to the penalty for each unit of downtime ($C_0 \Delta t_{delay}$) or to the rental cost of LRU paid to another airline ($C_{rent} \Delta t_{delay}$). C_0 is the penalty per unit of time for exceeding the duration of the warranty repair time and C_{rent} is the rent per unit of time for the use of LRU, leased from another airline.

It should be noted that the frequency of occurrence of the third situation is dependent on whether the supplier complies with the warranty repair duration t_{WR}. The greater the delay of the LRU recovery or replacement in the first and second situations, the higher the probability of occurrence of the third situation.

Let us denote the optimal number of spare LRUs in the warehouse of the airline by n_{opt}, calculated based on the values of t_{WR} and t_{ed}. Then, when exceeding the duration of the warranty repair or replacement of the LRU by the amount Δt_{delay}, the average time of aircraft delay is determined by the formula

$$\Delta T_{ad} = \Delta t_{LRU}\left(n_{opt}, t_{WR} + \Delta t_{delay}, t_{ed} + \Delta t_{delay}\right) + t_D + t_M - t_S \tag{18.28}$$

where Δt_{LRU} is the average time of delay in satisfying the demand for spare LRUs at the base airport, t_D and t_M are the average durations of dismantling and mounting the LRU on the aircraft board and t_S is the average duration of aircraft stop at the base airport while performing a typical route.

It is obvious that the following inequality holds:

$$\Delta t_{LRU}\left(n_{opt}, t_{WR} + \Delta t_{delay}, t_{ed} + \Delta t_{delay}\right) > \Delta t_{LRU}\left(n_{opt}, t_{WR}, t_{ed}\right) \tag{18.29}$$

Consequently, the probability of finding the avionic system in the state of waiting for a spare LRU at the aircraft stop is increased.

Equations for the calculation of $\Delta t_{LRU}(n_{opt}, t_{WR}, t_{ed})$ are given in [19].

The total average penalties for exceeding the duration of the warranty repair or replacement within the warranty period are determined as follows:

$$P(T) = \frac{mNT}{\tau} P_{delay} C_p \Delta T_{ad} \tag{18.30}$$

where P_{delay} is the probability that LRU failed in flight is awaiting a replacement during the aircraft stop at the base airport and C_p is the penalty (C_0) or the rent (C_{rent}) per unit of time.

The probability P_{delay} is given by

$$P_{delay} = \frac{\sigma\left[\Delta t_{LRU}\left(n_{opt}, t_{WR}, t_{ed}\right) + t_D + t_M - t_S\right]}{M\left(S_{op}\right) + M\left(S_{inop}\right) + \sigma\left[\Delta t_{LRU}\left(n_{opt}, t_{WR}, t_{ed}\right) + t_D + t_M - t_S\right]} \tag{18.31}$$

where $M(S_{op})$ and $M(S_{inop})$ are, respectively, the expected mean time spent by the LRU in the operable (S_{op}) and inoperable state (S_{inop}) and σ is the indicator function

$$\sigma = \begin{cases} 0 \text{ if } t_S \geq (\Delta t_{LRU} + t_D + t_M) \\ 1 \text{ if } t_S < (\Delta t_{LRU} + t_D + t_M) \end{cases} \tag{18.32}$$

Generalized equations for $M(S_{op})$ and $M(S_{inop})$ are given in [19]. In the case of exponential distribution of time to permanent and intermittent failure, $M(S_{op})$ and $M(S_{inop})$, are determined as follows [19]:

$$\begin{gathered} M\left(S_{op}\right) = \frac{\tau}{1 - e^{-\theta\tau}}\left[1 - e^{-(\lambda+\theta)T}\right] + \left[\left(1 - e^{-\lambda\tau}\right)\times\right. \\ \left.\left(\frac{1}{\lambda} - \frac{\tau}{1 - e^{-\theta\tau}}\right) - \tau e^{-\lambda\tau}\right]\frac{1 - e^{-(\lambda+\theta)T}}{1 - e^{-(\lambda+\theta)\tau}} + \tau e^{-(\lambda+\theta)T} \end{gathered} \tag{18.33}$$

$$M\left(S_{inop}\right) = \left(\tau - \frac{1 - e^{-\lambda\tau}}{\lambda}\right)\left[\frac{1 - e^{-(\lambda+\theta)T}}{1 - e^{-(\lambda+\theta)\tau}}\right] \tag{18.34}$$

As can be seen from (18.30)–(18.32), if $\sigma = 0$, then $P_{delay} = 0$ and $P(T) = 0$.

Example 2 Calculation of the expected penalty costs during warranty period for an avionic system if $T = 10{,}000$ h, $N = 5$, $m = 2$, $\tau = 5$ h, $\lambda = 10^{-4}$ h^{-1}, $\theta = 2 \times 10^{-4}$ h^{-1}, $t_{WR} = 120$ h, $t_{ed} = 16$ h, $C_0 =$ £20,000 and $t_M = t_D = 0.25$ h.

The dependence of $P(T)$ as a function of the number of spare LRUs is illustrated in Table 18.1. Several conclusions can be made from an analysis of Table 18.1. Firstly, it is obvious that the greater number of spare LRUs, the lower the expected penalty costs. Secondly, an increase of warranty repair duration ($2t_{WR}$) or expediting delivery time ($2t_{ed}$) results in a significant increase in the expected penalty costs. As can be seen from Table 18.1, the worst case is when both t_{WR} and t_{ed} are doubled.

Table 18.1 Calculated values of $P(T)$ for different values of t_{WR} and t_{ed}

Number of spare LRUs (n)	Expected penalty costs (£)			
	(t_{WR}, t_{ed})	($2t_{WR}$, t_{ed})	(t_{WR}, $2t_{ed}$)	($2t_{WR}$, $2t_{ed}$)
1	2.08×10^5	3.22×10^5	4.96×10^5	7.3×10^5
2	0	3.02×10^4	6.1×10^3	1.28×10^5
3	0	0	0	0

18.4.4 Capital Expenditures

Capital expenditures depend on whether the supplier is ready to install additional equipment for testing the dismantled LRUs. For example, the following two variants of warranty maintenance are evident. In the first variant, the restoration of all dismantled LRUs is carried out at the supplier factory. Here the capital expenditures are equal to zero. This variant can be cost-effective in the case of a small number of aircraft that have supplier warranty. The use of this variant when a large number of aircraft have supplier warranty requires an increase in the number of spare LRUs. In case of a shortage of spare LRUs, the expected penalty costs will increase according to (18.30). The second variant assumes that automatic test equipment (ATE) is used in the airline for re-testing the dismantled LRUs. The purpose of ATE is to avoid the shipment of falsely dismantled LRUs to the supplier for repairing.

Experience of modern aircraft operation confirms a rather high percentage of unscheduled removals of LRUs, which causes a significant increase in the number of spare LRUs needed to provide the required level of flight regularity. Furthermore, unscheduled LRU removals result in increased repair costs during the warranty period.

If the aircraft buyer does not purchase ATE, but it is profitable for the supplier to install ATE in the airline for rechecking dismantled LRUs, the capital expenditure recalculated to the beginning of operation of the first delivered aircraft is

$$C_{ATE,i}(1+\delta)^{i-1} \Big/ (1+\varepsilon)^{i-1}, \ i = 1, \ \ldots, T_0 \tag{18.35}$$

where T_0 is the duration of warranty expressed in calendar years, $C_{ATE,i}$ is the cost of ATE in the i-th year ($i = 1, \ldots, T_0$), δ denotes the increase of ATE cost due to inflation in labour costs and so on and ε is the depreciation in monetary value, since the expected cost of ATE would be affected by inflation, increased labour and other costs.

The current supplier costs in the j-th year of the interval $(0, T_0)$ associated with the renting of area to accommodate ATE in the airline, payment to staff, payment for electricity, and other necessary expenses is designated as C_j. Then, the capital expenditure for the organization of the warranty maintenance and operating costs during the interval $(0, T_0)$ is equal to:

$$C_E(T_0) = C_{ATE,i}(1+\delta)^{i-1} \Big/ (1+\varepsilon)^{i-1} + \sum_{j=i}^{T_0} C_j(1+\delta)^{j-1} \Big/ (1+\varepsilon)^{j-1} \tag{18.36}$$

If $i = 1$, i.e. the supplier installs ATE in the airline at the beginning of the warranty period, then (18.36) is converted to the form:

$$C_E(T_0) = C_{ATE,1} + \sum_{j=1}^{T_0} C_j(1+\delta)^{j-1} \Big/ (1+\varepsilon)^{j-1} \tag{18.37}$$

18.4.5 Cost of Purchasing an Excess of Spare LRUs

The supplier calculates the required number of spare LRUs in the airline's warehouse during the warranty period. The main parameters that determine the required number of spare LRUs are as follows: T—warranty period, N—number of aircraft that are under warranty by the supplier, λ—LRU permanent failure rate, θ—LRU intermittent failure rate, t_{WR}—warranty repair duration and t_{ed}—expedited delivery time of a spare LRU.

Let N^*, λ^*, θ^*, t_{WR}^* and t_{ed}^* be the values of the initial data used by the supplier for the calculation of the optimal number of spare LRUs, n^*. Since parameter N cannot be different from the specified N^*, the excess of spare LRUs may appear only as the difference between the values of λ^*, θ^*, t_{WR}^* and t_{ed}^* and the actual values of these parameters. For example, if the actual values of parameters λ and θ are less than λ^* and θ^*, there will be an excess of spare LRUs. Further, if the supplier repairs and replaces the failed LRUs within a time shorter than t_{WR}^* and t_{ed}^*, this will lead to an excess of spare LRUs.

Let n be the total number of spare LRUs purchased by the airline during the warranty period to ensure flight regularity. Then, the excess of spare LRUs to be re-purchased at the end of the warranty period is determined by

$$\Delta n = n^* - n \tag{18.38}$$

The expected costs for the purchase of the excess spare LRUs are determined by

$$C_{SP}(T_0) = \Delta n C_{LRU}(T_0) \Big/ (1+\varepsilon)^{T_0 - 1} \tag{18.39}$$

where $C_{LRU}(T_0)$ is the cost of LRU at the end of the warranty period.

Example 3 Calculation of expected costs for re-purchasing the excess of spare LRUs at the end of warranty period for an avionic system if $T_0 = 3$ years, $T = 10{,}000$ h, $N = 10$, $m = 3$, $C_{LRU} =$ £20,000, $\tau = 5$ h, $\lambda^* = 10^{-4}$ h^{-1}, $\theta^* = 2 \times 10^{-4}$ h^{-1}, $t_{WR}^* = 160$ h, $t_{ed}^* = 24$ h, $t_M = t_D = 0.25$ h and $\varepsilon = 0.1$.

For these initial data, the optimal number of spare LRUs n^* is 4. Now assume that some parameters had different values during the warranty period, for example, $\theta = 1 \times 10^{-4}$ h^{-1}, $t_{WR} = 120$ h and $t_{ed} = 16$ h. In this case, $n = 3$ and $\Delta n = 1$. Using (18.39), we calculate that $C_{SP}(T_0) =$ £16,530.

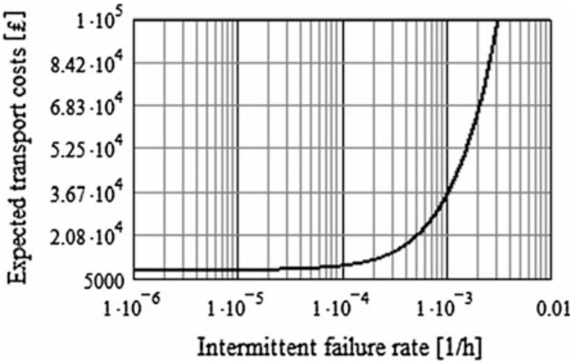


Fig. 18.3 Dependence of expected transport costs on intermittent failure rate

18.4.6 Expected Transport Costs for the Supplier

Avionics suppliers or buyers or both bear all transportation costs for shipping failed and restored LRUs. If transportation costs are fully borne by the buyer, the supplier expenses are zero. If the supplier pays for the transportation, fully or partially, the transport costs during the warranty period are determined analogous to (18.27)

$$C_{TC}(T) = \frac{mNTC_{tc}}{\tau}\left[P^*_{IF}(\tau) + P^*_{PF}(\tau)\right] \tag{18.40}$$

where C_{tc} is the transportation cost of shipping a failed LRU to the supplier and shipping the restored LRU back to the buyer.

Example 4 Calculation of expected transport costs for an avionic system, assuming T = 10,000 h, N = 10, m = 3, C_{tc} = £100, τ = 5 h and $\lambda = 10^{-4}\ \text{h}^{-1}$.

The dependence of expected transport costs as a function of intermittent failure rate is shown in Fig. 18.3. As can be seen from Fig. 18.3, $C_{TC}(T)$ begins to rise significantly when $\theta > 10^{-4}\ \text{h}^{-1}$.

If the supplier installed ATE at the buyer end, then only LRUs with permanent failures are directed to the supplier for repairs. In this case, (18.40) is simplified to

$$C_{TC}(T) = mNTC_{tc}P^*_{PF}(\tau)/\tau \tag{18.41}$$

18.5 Conclusions

In this chapter, we have analysed the components of the financial costs to avionic system suppliers during the warranty period. Mathematical models have been developed for the evaluation of the supplier's costs, taking into account the warranty duration, number of aircraft, reliability indicators with respect to permanent and intermittent failures, number of LRUs in a redundant avionic system, number of spare LRUs, cost of restoration and transportation and penalties for exceeding the duration of the warranty repair or replacement. These results enable a reasonable determination on the number of spare LRUs, minimize the amount of penalty costs in the case of breach of warranty by the supplier and minimize some other supplier costs. Practical use of the obtained results will significantly reduce the maintenance costs of avionics during the warranty period.

References

1. Aerospace warranty report. http://www.warrantyweek.com/archive/ww20160414.html Accessed: 14 April 2016
2. Blischke WR, Murthy DNP (1996) Product warranty handbook. Marcel Dekker Inc, New York
3. Thomas MU, Rao SS (1999) Warranty economic decision models: a summary and some suggested directions for future research. Oper Res 47(6):807–820
4. Chukova S, Hayakawa Y (2004) Warranty cost analysis: renewing warranty with non-zero repair time. Int J Reliab Qual Saf Eng 11(2):93–112
5. Jun B, Hoang P (2004) Discounted warranty cost of minimally repaired series systems. IEEE Trans Reliab 53(1):37–42
6. Huang HZ, Liu ZJ, Li Y et al (2008) A warranty cost model with intermittent and heterogeneous usage. Maintenance Reliab 4:9–14
7. Wang H (2006) Warranty cost models considering imperfect repair and preventive maintenance. IEEE Bell Labs Tech J 11(3):147–159
8. Park M (2010) Warranty cost analyses using quasi-renewal processes for multicomponent systems. IEEE Trans Syst Man Cybern Part A Syst Hum 40(6):1329–1340
9. Blischke WR, Karim MR, Murthy DNP (2011) Warranty data collection and analysis. Springer series in reliability engineering
10. Diaz VG, Campos ML, Fernandez JFG (2010) Warranty cost models state-of-art: a practical review to the framework of warranty cost management. In: Bris R, Soares G, Martorell R (eds) Reliability risk and safety: theory and applications, pp 2051–2059
11. Shafiee M, Chukova S (2013) Maintenance models in warranty: a literature review. Eur J Oper Res 229(3):561–572
12. Bieber P, Boniol F, Boyer M et al (2012) New challenges for future avionic architectures. Aerosp Lab J 4:1–10
13. Wang Y, Song B (2008) Manpower management benefits predictor method for aircraft two level maintenance concept. Modern Appl Sci 2(4):33–37
14. National Security and International Affairs Division (1996) Two level maintenance program assessment, United States general accounting office. AD-A307070
15. Raza A, Ulansky V (2015) Minimizing total lifecycle expected costs of digital avionics' maintenance. In: Proceedings of 4th international through-life engineering services conference, Cranfield, Nov 2015. Procedia CIRP, vol 38, pp 118–123
16. Drenick RF (1960) The failure law of complex equipment. J Soc Ind Appl Math 8(4):680–690
17. Salemi S, Yang L, Dai J et al (2008) Physics-of-failure based handbook of microelectronic systems. University of Maryland

18. Qin J, Huang B, Walter J et al (2005) Reliability analysis in the commercial aerospace industry. J Reliab Anal Center Dep Def USA 1:1–5
19. Raza A, Ulansky V (2016) Assessing the impact of intermittent failures on the cost of digital avionics' maintenance. In: Proceedings of IEEE aerospace conference, Big Sky, Montana, pp 1–16

Chapter 19
Product-Service Systems Under Availability-Based Contracts: Maintenance Optimization and Concurrent System and Contract Design

Amir Kashani Pour, Navid Goudarzi, Xin Lei and Peter Sandborn

Abstract Product-service systems (PSSs) are the result of a shifting business focus from designing and selling physical products, to selling a system consisting of products and services in an ongoing relationship with the customer that fulfills customer satisfaction. A PSS contract can take several forms (e.g., fixed price, capability-contract, and availability-based). The focus of this chapter is on PSSs that use availability-based contracts. In these cases the customer does not purchase the product, instead they purchase the utility of the product and the availability of service in order to obtain a lower cost while still meeting their needs. This chapter addresses the optimization of system maintenance activities, and the concurrent design of the PSS and the contract.

19.1 Introduction

Performance-based contracts (PBCs) and similar mechanisms have become popular for contracting the sustainment of military systems in the United States and Europe. Performance-based logistics (PBL), also referred to as performance-based life-cycle product support and contracting for availability, refers to a group of strategies for system support that instead of contracting for goods and services, a contractor delivers performance outcomes as defined by performance metric(s) for a system or product [1]. Performance-based thinking is reflected in a famous quote from Levitt

---

A. Kashani Pour · X. Lei · P. Sandborn (✉)
Center for Advanced Life Cycle Engineering, University of Maryland,
College Park, MD 20742, USA
e-mail: sandborn@umd.edu

N. Goudarzi
Mechanical Engineering Department, University of North Carolina Charlotte,
Charlotte, NC 28223, USA

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_19

[2]: “The customer doesn’t want a drilling machine; he wants a hole-in-the-wall.” PBL and similar outcome-based contracts pay for effectiveness (availability, readiness and/or other performance measures) at a fixed rate, penalize performance shortcomings, and/or award gains beyond target goals.

The impact of a performance-based contract oriented design process on original equipment manufacturer (OEM) decision making for optimizing reliability in the post-production purchase period led to the development of integrated schemes with dynamic interdependencies of the product and the service the product provides called product-service systems (PSSs) [3].[1] One example of a PSS is the function of washing clothes using a washing service. In this case, customers pay for the laundering of their clothes instead of buying the washing machine. This example, called pay-per-wash, is described in [4]. When providing a function instead of a product, a contract must be entered into between the customer and the service provider (the contractor). Here, the connection between the stakeholders becomes formal, and the contracts that regulate what the offer includes are important [5].

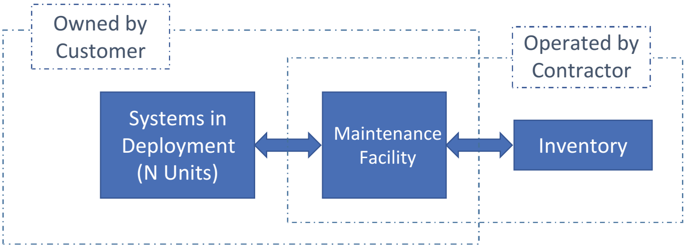
Procurement and system acquisition process efficiency and success across a system’s life cycle requires the development and implementation of best-value, long-term, performance-based product support strategies that leverage performance-based agreements with both industry and government product support providers [6]. Hence, an effective combination of technical and monetary approaches that includes the inventory, maintenance, and operational decisions together to form a unified model that provides visibility into the effect of different decisions is required [7]. Performance-based logistics (PBL) contracting is intended to incentivize this integration towards concurrently reducing life-cycle cost and improving performance.

Availability-based contracts that pay for maintenance or service effectiveness and penalize shortcomings in the specific availability metric (e.g., materiel, operational, or inventory-level availability) have been introduced to incentivize cost reduction efforts and guarantee readiness on the contractor side of PSS. These types of contracts are also referred to as “performance-based” contracts (PBC), “outcome-based contracting maintenance models,” “performance-based logistics” (PBL) and “contracting for availability” (CfA). Availability-based contracting concepts are being used for PSS acquisitions in healthcare, energy, military systems and infrastructure. These contracts allow customers to pay only for the specific outcomes (e.g., availability) achieved rather than the workmanship and materials being delivered.

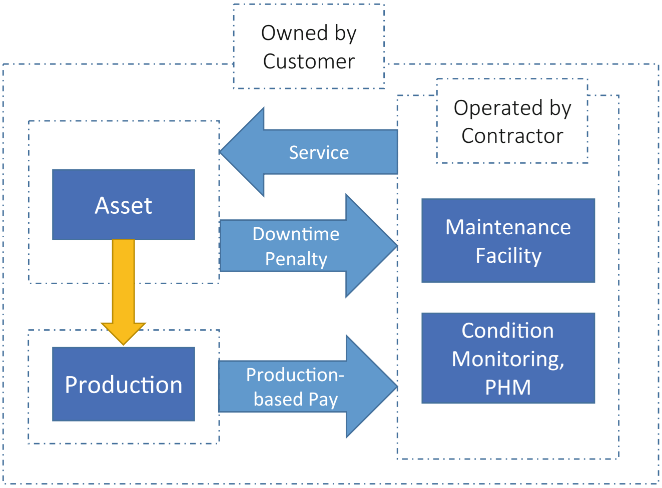
Every PSS has two sides: the customer who expects a specific level of outcome (e.g., availability) over the period of the contract, and the contractor who provides the outcome for the period of the contract and its possible extension.

---

[1] We assume that ideally the design of the PSS means designing the hardware, software, service, and logistics associated with the system concurrently. Section 19.4 of this chapter includes contract design in this process as well.



(a) Non-revenue earning for the customer



(b) Revenue earning for the customer

Fig. 19.1 Repairable item flow under an outcome-based contract describing possible customer/contractor relationships

There are several possible ways to share the tasks, risks and measure the outcome in these contracts. Figure 19.1 illustrates two possible contractor/customer relationships. For example, Fig. 19.1a shows a case where the contractor operates a repair facility that is owned by the customer and is in charge of replacement and repair of the systems. In this case, the contractor is only committed to maintain the inventory availability (e.g., [8]) or is in charge of the availability of the deployed systems using fleet-level availability metrics [9].[2] The contractor incurs the costs associated with shipping, holding and ordering, and receives a payment from the customer. The customer generally does not produce any income from the operation

---

[2] Classically "inventory" refers to an inventory of items (e.g., spare parts), however, more generally, it could mean a maintenance "opportunity inventory", which is a combination of all the resources necessary to support the system, i.e., workforce, facilities, favorable weather, and spare parts. This broader interpretation of inventory (previously eluded to by [10]) is a departure from mainstream operations research that only thinks of inventory as parts.

and one can call this process non-revenue earning from the customer's perspective. The optimization of logistics, maintenance and the system's design will be done by the contractor without the customer's involvement.

Another possible contractor/customer relationship that is common for systems that are revenue-earning for the customer is shown in Fig. 19.1b. In this case the relationship is based on two-levels of production payment and downtime penalty based on opportunity cost. The customer makes revenue from production and pays the contractor based on the combination of its revenue as well as penalizing for operational downtime. In this perspective the availability of production is the critical attribute for the customer, but the measure of production is also dependent on other production factors (e.g., in a wind farm it depends on wind speed). From the contractor point of view there is a payment or penalty for every hour of operation.

Revenue-earning and non-revenue-earning are customer distinctions, from the contractor's viewpoint, everything is revenue earning (if it wasn't there would be no contract). Systems can also be distinguished based on the form of the outcome. For production systems the contractor's compensation is determined by a payment schedule that is based on the amount or quantity of outcome the system produces. For non-production systems, the contractor's compensation is determined by a payment schedule that is based on the availability of the system.

One of the merits of availability-based contracts is the optimal sharing of risks by both parties. However, this presents several challenges. First, the optimization of system sustainment differs depending on whether: (a) the system is managed in isolation, or (b) the system is part of a larger population of systems managed by an availability-based contract. Secondly, availability-based contracts present a pricing challenge due to the stochastic nature of system's performance, customer usage, the outcomes of the contractor's decisions, and the impact of these contracts on the risk-sharing balance. Failures to understand the PSS's sustainment impacts and the risks involved, and therefore the contract cost, have caused some projects to stop and given rise to doubts about the applicability of this class of strategies for new acquisition contracts.

This chapter has two parts. First we discuss the incorporation of availability-based contract requirements into the optimal design of the maintenance of systems (Sect. 19.3), and second, progress toward concurrent PSS design and contract design (Sect. 19.4). However, before discussing the maintenance optimization and concurrent design aspects, it is useful to briefly review general contract modeling.

19.2 Contract Modeling

Contracts can be modeled from several different viewpoints, however, it is most useful to model the class of contracts discussed in this chapter from the contractor's point of view as a payment model that addresses the desired outcome constraints in which performance and cost are weighted into a single factor to simplify the payment rationale [9],

$$P_t(C_t, A_t) = \omega_t + \alpha_t C_t + v_t A_t \tag{19.1}$$

where P_t is the payment to the contractor, C_t is the cost to the contractor, A_t is the availability, and $\omega_{t,}$ α_t and v_t are the contract parameters chosen by the customer and described in the contractual document where t refers to the monitoring interval. Note, for simplicity, Eq. (19.1) ignores the contractor's profit. ω_t is the fixed payment. α_t is the customers' share of the contractor's costs of operation and v_t is the penalty or award rate for achieved availability below or above the required level.

By varying the contract parameters in Eq. (19.1), different classes of cost-driven and outcome-based contracts can be modeled. For example, $v_t = 0$ and $\alpha_t = 0$ is a fixed-price contract, and $v_t = 0$ and $\alpha_t = 1$ is a cost-plus contract with full reimbursement. Since this model is completely known to the contractor it is safe to assume that they optimize their decisions based on the above model. However, their decisions might incur costs on the customer outside of the scope of the contract, for example maintenance costs after the contract is over (i.e., so called silent hazards).

From the viewpoint of customer these contracts can be modeled by a Stackleberg game in which, depending on the contract designed by the customer, the contractor will optimize its strategy. From this point of view a modeling can be done by a two-level optimization problem [11].

In this chapter we consider two different ways of modeling contracts. In Sect. 19.3 we use a payment model to directly model the contractor and we ignore the interaction between the contractor and the customer except when considering discrete maintenance opportunities. In Sect. 19.4.2 we describe a two-level stochastic optimization schemed used to optimize the contract and the maintenance actions.

19.3 Optimization of Maintenance Activities in PSSs Under Availability-Based Contracts

This section presents the concept of PHM-enabled maintenance options. Then, it describes how the requirements from an availability-based contract are incorporated into an option valuation process in order to optimize the maintenance planning for systems.

19.3.1 System Health Management as a Maintenance Design Activity

The maintenance planning that this chapter focuses on is contingent on the presence and use of system health management technologies. System health management technologies such as condition-based maintenance (CBM) seek to perform predictive maintenance based on the condition of the system. Prognostics and health management (PHM) uses the condition of the system coupled with the expected future environmental conditions (temperature, vibration, etc.) to forecast a remaining useful life (RUL). The system management challenge is how to perform an accurate system risk allocation using the predicted RULs (with their associated uncertainties) to optimally plan when to perform predictive maintenance and allocate maintenance resources. The optimal maintenance planning is modified by performance requirements imposed by the availability-based contracts.

19.3.2 Maintenance Planning Using Real-Option Analysis

This section presents the concept of PHM-enabled predictive maintenance options. In Sect. 19.3.3 we describe how the requirements from an outcome-based contract are incorporated into the option valuation process.

A real option is the right, but not the obligation, to undertake certain business initiatives, such as deferring, abandoning, expanding, staging, or contracting. For example, the opportunity to invest in an asset is a real “call” option. Real options differ from financial options in that they are not typically traded as securities, and do not usually involve decisions on an underlying asset that is traded as a financial security. Unlike conventional net present value analysis (discounted cash flow analysis) and decision tree analysis, real options model the flexibility management has to alter the course of action in a real asset decision, depending on future developments. Predictive maintenance options are created when in situ health management (i.e., PHM) is added to systems. In this case the health management approach generates a remaining useful life (RUL) estimate that can be used to take proactive actions prior to the failure of a system. The maintenance option when PHM is used is defined by Haddad et al. [12],

- Buying the option = paying to add PHM to the system
- Exercising the option = performing predictive maintenance prior to system failure after an RUL indication
- Exercise price = predictive maintenance cost
- Letting the option expire = doing nothing and running the system to failure then performing corrective maintenance

The value from exercising the option is the sum of the cumulative revenue loss and the avoided corrective maintenance cost (corrective maintenance being more expensive than predictive maintenance).

The cumulative revenue loss is what the system would earn between the predictive maintenance event and the end of the RUL (if no predictive maintenance was done). Restated, this is the portion of the system’s RUL that is thrown away when predictive maintenance is done prior to the end of the RUL. In reality, this cumulative revenue takes the form of loss in spare part inventory life (i.e., the revenue earning time for the system will be shorter because some inventory life has been disposed of) [13].

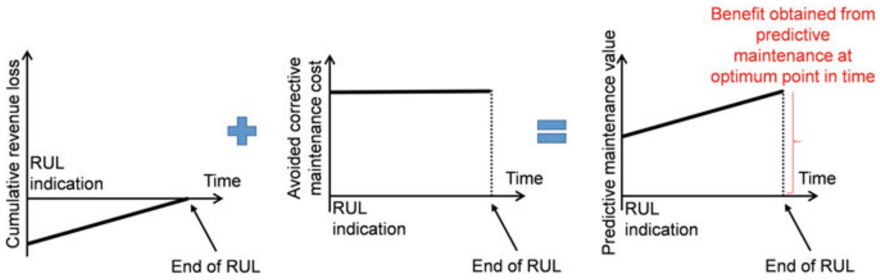


Fig. 19.2 Predictive maintenance value construction [13]

Avoided corrective maintenance cost includes[3]: the avoided corrective maintenance parts, service and labor cost, the avoided revenue loss associated with corrective maintenance downtime and the avoided under-delivery penalty due to corrective maintenance (if any).

Figure 19.2 illustrates the construction of the predictive maintenance value. The cumulative revenue[4] loss is the largest on day 0 (the day the RUL is forecasted). This is because the most remaining life in the system is disposed of if predictive maintenance is performed the day that the RUL is predicted. As time advances, less RUL is thrown away (and less revenue is lost). The avoided corrective maintenance cost is assumed to be constant. The predictive maintenance value is the summation of the cumulative revenue loss and the avoided corrective maintenance cost (Fig. 19.2). If there were no uncertainties, the optimum point in time to perform maintenance would be at the peak value point (at the RUL), which is the last moment before the system fails. Unfortunately, everything is uncertain.

The primary uncertainty is in the RUL prediction. The RUL is uncertain due to inexact prediction capabilities, and uncertainties in the environmental stresses that drive the rate at which the RUL is used up. A "path" represents one possible way that the future could occur starting at the RUL indication (Day 0). The cumulative revenue loss paths have variations due to uncertainties in the system's availability or uncertainties in how compensation is received for the system's outcome.[5] The avoided corrective maintenance cost paths represent how the RUL is used up and vary due to uncertainties in the predicted RUL. Each path is a single member of a population of paths representing a set of possible ways the future of the system could play out.

---

[3] This is not the difference between the predictive and corrective maintenance actions, but rather the cost of just a corrective maintenance event. The predictive maintenance event cost is subtracted later when the real option value is determined, i.e., in Eq. (19.2).

[4] The value construction in this section assumes that the system is revenue earning, e.g., a wind turbine or an airplane used by an airline.

[5] For example, if the system is a wind turbine, path uncertainties could be due to variations in the wind over time.

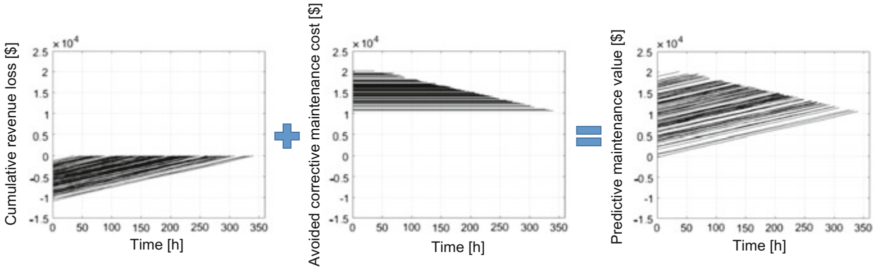


Fig. 19.3 Example of the simulated paths after an RUL indication

Due to the uncertainties described above, there are many paths that the system can follow after an RUL indication as shown in Fig. 19.3. Real options analysis lets us evaluate the set of possible paths to determine the optimum action to take.

Consider the case where predictive maintenance can only be performed on specific dates.[6] On each possible maintenance date, the decision-maker has the flexibility to determine whether to implement the predictive maintenance (exercise the option) or not (let the system run to failure, i.e., let the option expire)[7]. This makes the option a sequence of "European" options that can only be exercised at specific points in time in the future. The left side of Fig. 19.4 shows two example predictive maintenance paths (diagonal lines) and the predictive maintenance cost (the cost of performing the predictive maintenance). Real Option Analysis (ROA) is performed to valuate the option where the predictive maintenance option value is given by

$$O_{PM} = Max(V_{PM} - C_{PM}, 0) \tag{19.2}$$

where V_{PM} is the value of the path (right most graph in Fig. 19.3 and the diagonal lines in Fig. 19.4), and C_{PM} is the predictive maintenance cost. The values of O_{PM} calculated for the two example paths shown on the left side of Fig. 19.4 are shown on the right side of Fig. 19.4. Note that there are only values of O_{PM} plotted at the maintenance opportunities (not between the maintenance opportunities).

---

[6] This could be due to the limited availability of maintenance resources or the limited availability of the system being maintained.

[7] The decision-maker may also have the flexibility not to implement the predictive maintenance on a particular date but to wait until the next possible date to decide, which makes the problem an American option style as has been demonstrated and solved by [12]. The Haddad et al. solution in [12] is correct for the assumption that an optimal decision will be made on or before some maximum waiting duration and the solution delivered is the maximum "wait to date". Unfortunately, in reality maintenance decision-makers for critical systems face a somewhat different problem: given that the maintenance opportunity calendar is known when the RUL indication is obtained, on what date should the predictive maintenance be done to get the maximum option value. This makes the problem a European option style.

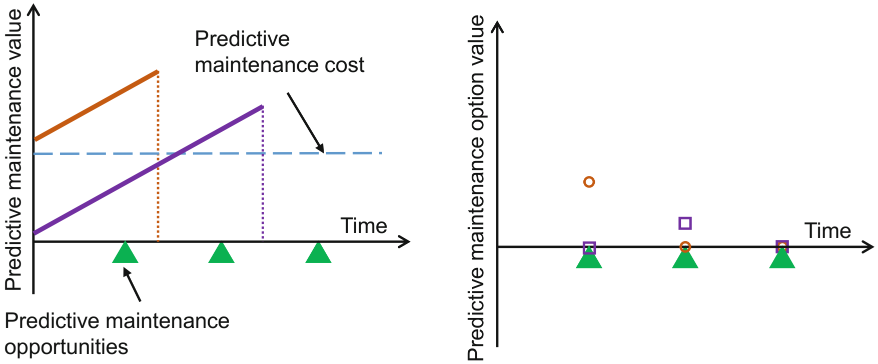


Fig. 19.4 Real options analysis (ROA) valuation approach. *Right graph*: *circles* correspond to the upper path and the *squares* correspond to the lower path in the *left graph*. The *triangles* indicate predictive maintenance opportunities

Equation (19.2) only produces a non-zero value if the path is above the predictive maintenance cost, i.e., the path is "in the money".

Each separate maintenance opportunity date is treated as a European option. The results at each separate maintenance opportunity are averaged to get the expected predictive maintenance option value of a European option expiring on that date. This process is repeated for all maintenance opportunity dates. The optimum predictive maintenance date is determined as the one with the maximum expected option value. The detailed mathematical formulation of the solution can be found in [13].

19.3.3 *Incorporating Availability-Based Contract Requirements into the Predictive Maintenance Option*

The "paths" described in Figs. 19.2 and 19.3 are based on a production system with an "as-delivered" energy delivery contract for a wind farm that defines a single fixed price for each unit of the energy delivered. When a system is managed via an availability-based contract (like a PBL), the paths shown in Fig. 19.3 will be impacted. The availability-based contract influences the combined predictive maintenance value paths due to changes in the cumulative revenue loss and the avoided corrective maintenance cost paths. These paths will be influenced by the availability target, prices before and after that target is reached (generally the latter is lower than the former), penalization mechanisms, the availability already produced, and the operational state of the other systems in the population. For example, assume that the cumulative availability produced by a population of systems is close to the availability target. All systems are operational while some are indicating

RULs. The population of system can meet the availability target without the members indicating RULs. Then the cumulative revenue loss of the systems with RULs will be lower than when they are managed under a non-availability-based contract, since the cumulative revenue loss will be lower (because the price paid for the availability is lower after the outcome target is met). Assume a different scenario where the cumulative outcome from the population of systems is far from the outcome target, and many systems are non-operational. In this case, running the systems with RULs to failure and performing corrective maintenance causing long downtimes may result in the population of the systems not reaching the outcome target. In this case the under-delivery penalty will occur, and the avoided corrective maintenance cost will be higher than the non-availability-based contract (as delivered) case that doesn't have any penalization mechanisms.

Under an availability-based contract, the optimum predictive maintenance opportunity for individual systems in a population (e.g., a fleet) are generally different than for an individual system managed in isolation. These two cases would have the same optimum if an as-delivered contract was used.

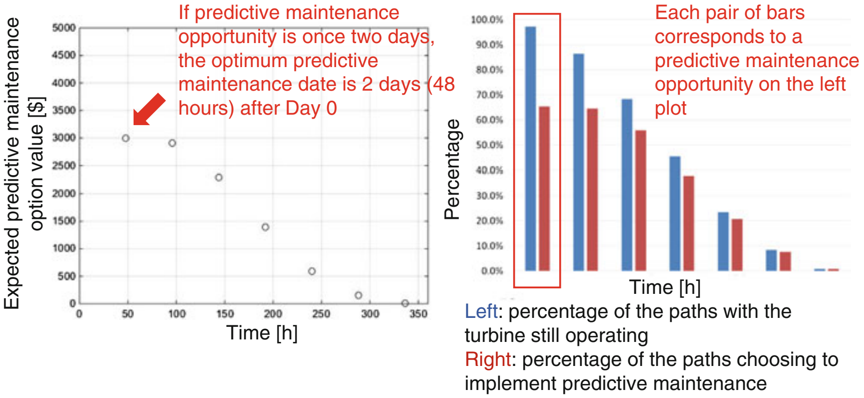
19.3.4 A Wind Turbine with an Availability-Based Contract

In this Section, the predictive maintenance option model is implemented on a single turbine and then a wind farm with multiple turbines is managed via an availability-based contract. A Vestas V-112 3.0 MW offshore wind turbine [14] was used in this example.

Maintaining offshore wind turbines requires resources that are not continuously available. These resources include ships with cranes, helicopters, and trained maintenance personnel. These resources are often onshore-based (which may be as much as 100 miles from the wind farm) and may be maintaining more than one wind farm. Therefore, maintenance is only available on scheduled dates (maintenance opportunities) that may be weeks apart. The availability of maintenance is also dependent on weather and ocean conditions making the timing of future maintenance visits uncertain.

Figure 19.5 shows an example result for a single wind turbine. In this example, the ROA approach is not trying to avoid corrective maintenance, but rather to maximize the predictive maintenance option value. In this example, at the determined optimum maintenance date the predictive maintenance will be implemented on only 65.3% of the paths (the paths that are "in the money"). 32.0% of the paths, which are "out of money", will choose not to implement predictive maintenance, and in 2.7% of the paths the turbine has already failed prior to that date.

The result in Fig. 19.5 assumes that all the power generated by the turbine can be sold at a fixed price. There are many wind farms (and other renewable energy power production facilities) that are managed under availability-based contracts called power purchase agreements (PPAs). A PPA defines the energy delivery targets, purchasing prices, output guarantees, etc. Wind farms are typically



If predictive maintenance opportunity is once two days, the optimum predictive maintenance date is 2 days (48 hours) after Day 0

Each pair of bars corresponds to a predictive maintenance opportunity on the left plot

Left: percentage of the paths with the turbine still operating
Right: percentage of the paths choosing to implement predictive maintenance

Fig. 19.5 Optimum maintenance date after an RUL indication for a single wind turbine

managed via PPAs for several reasons [15]. First, though wind power can be sold into the spot market, the average spot market prices vary greatly (above and below) the long-term PPA contract prices. Second, lenders are not willing to finance wind projects without a signed PPA that secures a future revenue stream. Third, wind energy buyers prefer simply purchasing power to building and operating their own wind farms.

PPA terms are typically 20 years for wind energy, with either a constant or escalating contract price defined through the whole term. At the beginning of each year, a PPA often requires the seller to estimate how much energy they expect to generate during the whole year, based on which an annual energy delivery target. For energy that is beyond the annual energy delivery target a lower excess price may apply. The buyer may also have the right not to accept the excess amount of energy, or adjust the annual target of the next contract year downward based on how much has been over-delivered. A minimum annual energy delivery limit or output guarantee may also be set, together with a mechanism to determine the liquidated damages. For example, the seller must compensate the buyer for the output shortfall that the buyer is contracted to receive, multiplied by the difference between the replacement energy price, the price of the energy from sources other than wind paid by the buyers to fulfill their demands, and the contract price. The buyer may also adjust the annual target of the next contract year upward to compensate for how much has been under-delivered.

Assume a 5-turbine-farm managed via a PPA, Turbines 1 and 2 indicate RULs on Day 0, turbine 3 operates normally, and turbines 4 and 5 are non-operational. Predictive maintenance value paths of all turbines with RULs need to be combined together because maintenance will be performed on multiple turbines on each visit (see [16] for details on how the paths are combined for multiple turbines). Cumulative revenue loss, avoided corrective maintenance cost, and predictive maintenance value paths for turbines 1 and 2 are shown in Fig. 19.6.

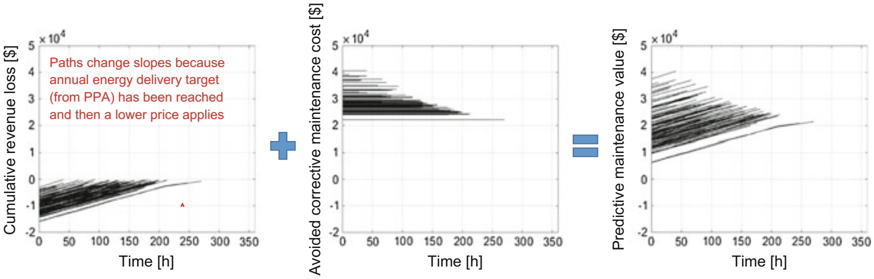


Fig. 19.6 Combined value paths for turbines 1 and 2 in a 5 turbine farm managed by a PPA. Some paths (as indicated by the *arrow*) change slopes because the annual energy delivery target defined by the PPA has been reached after which a lower price for the power applies

19.4 Concurrent PSS and Contract Design

In Sect. 19.3, the PPA (the contract) sets the requirements for the system, and the maintenance planning is optimized based on those requirements. This is a case where the availability-based contract design has been performed separately from the through-life engineering services (TES) and engineering design processes, and provided as a requirement to the engineering design process. This approach may create significant risks for both parties. For systems that are subject to availability-based contracts, contract failure may mean large amounts of money spent by the customer (potentially the public) for either no availability or inadequate availability, or result in the contractor being driven out of business, which can lead to disaster for both parties.

A fundamental gap exists between contract design and engineering (PSS) design. This section discusses the current state of contract/PSS design and progress toward concurrent system and contract design.

19.4.1 State-of-the-Art in Contract-Based System Design

Traditionally, the contract and product parameters are defined separately. In recent years, driven by a need for enhancing system reliability, maintainability, and logistics support, attempts to include contract and engineering (performance) parameters simultaneously have been articulated, but has not been done. There are a significant number of papers with a wide array of measures to determine performance, taking both objective and subjective views.

In this section, the relevant approaches for designing contracts and products are reviewed and the need for concurrent contract-engineering design is introduced as a key solution to obtain a more realistic overall PSS design.

The correlation between contracts and the PSS design process can be classified into three categories:

(1) Engineering/logistics design using fixed contract parameters

In this category, it is assumed that the contract parameters are given as a set of requirements, and they are treated as fixed input parameters in the PSS design (i.e., they are constraints on the PSS design). Hence, the PSS parameters are designed to maximize the operating performance and functionality that satisfies the contract requirements.

Examples of product design processes (hardware and/or software) that include one or more contract parameters, e.g., cost constraints, length of support requirements, etc., are very common. The analysis in Sect. 19.3 is an example of this category of work where PPA requirements (energy price and the annual delivery target are used to perform maintenance planning design for the wind farm). Other examples include Nowicki et al. [17] who developed a spare provisioning system to respond to a given performance-based contract from the viewpoint of the contractor. In [17] the contractor's objective is to maximize profit and the scope of its activity by optimizing the inventory level (the inventory level is considered to be part of the logistics design). This scheme also includes sensitivity analysis that addresses the reliability of the product.

Less common are PSS design processes that use actual availability requirements. Jazouli et al. [18, 19] estimated the required logistics, design, and operation parameters for a specific availability requirement. In this work the developed model connects the requirements on each operational decision regarding repair, replacement and inventory lead-time so that the impact of the contract terms can be seen on the logistics decisions. Jin and Wang [20] studied the impact of reliability and usage uncertainty on planning PBCs incorporating equipment availability, mean-time-to-failure, and mean-time-to-repair.

(2) Contract design that uses fixed product parameters

In this category, the contract parameters are optimized for a given PSS. For example, the following contract parameters may be determined: the payment schedules (amount and timing) [21], profit sharing [22], the length of contract [23], the selected contract mechanism [17, 24], supply-chain parameters (inventory lead time,[8] back-order penalties, etc.) [25], and warranty[9] design could be determined.

---

[8] In [18] the inventory lead time (ILT) was considered to be a logistics parameter determined from an availability requirement. It is also possible that ILT is a contract parameter that is flowed down to subcontractors.

[9] Although we include warranty design in the list of possible contract design activities that could be driven by the product parameters, for most products that have warranties, the type of warranty and its length are determined by marketing, and are not based on the product's predicted reliability. More commonly, the warranty type and length (which are a contract) are passed to the engineering design to determine the appropriate warranty reserve fund, which would be an example of the first category.

Examples of work in this category include Arora et al. [26] who studied an integrated inventory and logistics model to minimize the total cost of supply-chain support. Nowicki et al. [17] developed a model that designs performance-based contracts with different lengths and contract fees. In this work the contract design is based on a given product with a fixed initial reliability. They explore the opportunity for further investment in improvements in the product's reliability under the proposed PBC to demonstrate a win-win for the customer and contractor through the optimal choice of contact length.

Hong et al. [24] employed mechanism design theory[10] to design an optimized maintenance service contract for gas turbines in which uncertainties associated with customer actions, engine performance, and maintenance costs during the contract execution phase were accounted for. They assumed that the gas turbine design was given and determined the contract that maximizes the expected profit and provides a win-win incentive for the customer and contractor.

Wang [27] developed and discussed three different contract options for maintenance service contracts between a customer and a contractor for a given system design. The contract options were: (1) a full contract that covers both inspections and inspection repairs, and failure repairs, (2) a partial contract that covers inspections and inspection repairs, but not the failure repairs, and (3) a partial contract that covers failure repairs only.

In this category, there are several challenges. The existing models require a better understanding of the impact of incentive structures on the system design and usage. Zhu and Fung [25] proposed a model based on the service delivery and customer satisfaction level. They studied the design of optimal contracts that balance the incentives and risks on the two sides of a contract, so that both can achieve maximum profits. They assume that incentive payments to the contractor are dependent on the contractor's performance. Further research is also required on the risk attitude of contractors: risk-aggressive, risk-averse, or prudent. In addition, a more general and comprehensive model would include flexibility for the service provider to change their level of effort during the project to increase the chances of meeting their contractual goals. Moreover, an important gap in contract theory models is the assumption of a static risk allocation for the entire length of a project.[11] Zhao and Yin [28] propose a theoretical model for a dynamic risk allocation in constructing a project. However, a successful dynamic risk allocation needs a comprehensive understanding of both engineering and contractual parameters and their variations throughout a project. Such a dynamic risk allocation is not addressed in any theoretical models and is the subject of the next category.

---

[10] Mechanism design theory is an economic theory that seeks to determine when a particular strategy or contract mechanism will work efficiently.

[11] This problem is also evidenced by the choice of a single value for the cost of money, i.e., the cost of money is not constant over time (nor the same for all projects within an organization).

(3) Concurrent design of the contract and the PSS

Finally, the concurrent design of both the contract and the PSS would be the ideal solution (for both the customer and contractor) for real applications. However, there are no models that accurately assess and design CfA, dealing with all the risks and uncertainties involved [29]. One important proposed solution to fill this gap is to use engineering inputs and to find the engineering connections to current theoretical contract models [30]. Kashani Pour et al. [31] reviewed existing analytical models in this space and developed a framework for the design of availability-based contracts with consideration of engineering design and incentive structure.

There is an increasing interest in employing PBC concepts to obtain a better mutual understanding between the supplier and the customer. However, the existing literature is primarily focused on solving the problem from the contractor point of view and does not address the role of optimum contract design from the customer's viewpoint. This is partially due to the relatively short history of this class of contract [29], a lack of sufficient public data on different design contracts, and ignorance of the dynamic impact of uncertainties in the existing models.

A few authors discuss the need for concurrent design, e.g., [17]; even fewer attempt to provide any type of solution to the problem [24], and in cases that claim to address both the customer and contractor, the solutions are primarily sensitivity analyses that ignore the asymmetry of information or moral hazard problem.[12] Another proposed approach (also sensitivity analysis) is to study the impact of engineering parameters on the construction of contracts [32]. Sols et al. [32] studied the formulation of an n-dimensional performance-based reward model for use in PBC contracts. They developed an n-dimensional metrics structure that represents the system effectiveness along with its reward model that results in a successful PBC contract.

The type of cost modeling necessary for concurrent engineering and contract design isn't the same as for either engineering or contract design alone [5]. Most of the current CfA decisions are based on expert opinions, estimation, and historical data from previous designs, which can be unreliable [33]. In addition, such an approach is less useful when system complexity increases [34]. Also, a lack of relevant historical data is a major source of challenge in new projects [33, 34].

Based on Kashani Pour et al. [31], solutions provided in this category should be able to address the requirements breakdown (or flow down) to subcontractors. The breakdown of requirements for use by sub-system designers shares the freedom provided by availability-based contracts. Solutions are also required to provide concise algorithms so that the availability will be tangible

---

[12] While there are some major manufacturers who appear to (or claim to) use an integrated approach in concurrently designing contract and product parameters, they are unpublished and no details are available.

and measureable, and so the contractor can implement and understand the requirements within their sustainment activities. Designing availability-based contracts should address reliability design of products and operational decisions based on condition monitoring technologies, the role of incentives and their impact on the life-cycle of the product, supply-chain management of the PSS, and the integrated design and joint optimization of different performance metrics. These requirements make the use of concurrent design of PSS and contracts a necessary approach to model the problem for application to real-world practice.

The key questions that should be answered in this category are: (1) What are the main elements of an availability-based contract for a PSS? (2) What are the essential attributes of the concurrent PSS and contract design process? And (3) How are the advantages of concurrent design of PSS and contracts versus the first two category of design verified?

To summarize the concurrent design of the contract and PSS needs to address both the contractor and customer and the dynamics created by the contractual term between them including addressing uncertainties in achieving availability or reliability-related challenges. Concurrent design considers contract design as a part of integrated system design with PSS and the contract of main sub-systems with a dynamic relationship that is subject to stochastic processes such as reliability, supply-chain demand and operational uncertainties.

19.4.2 Contract Design as a System Design Problem

We approach contract design concurrent with the PSS as a system design problem where the process of designing contractual terms that address performance metrics, the payment model, and performance assessment are design parameters and a multidisciplinary life-cycle simulation of design impacts needs to be integrated into the engineering design process. The significant challenge of contract design in practice is on the customer side.

In the case of availability-based contracts, the TES and engineering designs should determine the contract requirements and the contract length and price in the acquisition and procurement stage, so that it protects the interests of the customer throughout the life cycle (i.e., it does not overpay the contractor, but also minimizes the risk that the system will become unsupported).

19.4.2.1 A Concurrent Contract and System Design Example

The study summarized in this section is the availability-based contract for spare-part availability for a submarine fleet (in this case the spare part is a torpedo). Each item in the inventory is a spare subassembly for a fleet of end-products that are carrying out a specific mission. Exercises and deployment requires continuous service of torpedoes before being returned to the fleet for usage. When the torpedoes are being tested, if they are found to be defective, they are replaced from inventory by supply contractors. The data and details of this example can be found in [7]. In this case the relationship between contractor and customer is similar to Fig. 19.1a.

The objective of the contract design process is to determine the best penalty and reward rate along with the best assessment interval (the interval over which to measure availability) as a contract parameter to reduce the total cost of the system and guarantee availability requirements. This optimum assessment interval should also be robust with respect to uncertainties in the costs that the contractor will incur. Since the cost model of the contractor is determined by penalty function parameters such as back-order and holding cost, and these costs are not related to the customer, a careful study of these contractual parameters needs to be done by the customer.

Back-order cost is primarily driven by opportunity cost, the cost of down-time or in general unavailability cost. It can be given a fixed pre-determined value or be connected to a variable source similar to production-guarantee contracts. Knowing all these helps the customer to decide on the cost of a back-order.

The contractor's cost of holding an inventory comes from the actual cost of operating the inventory facility and the penalties that are incurred because the inventory level is kept low. In this case the inventory facility is owned by the customer and the customer defines the holding costs that the contractor pays; in general the customer prefers a higher reliability and a lower level of inventory. The penalty is assigned to a actual holding cost (h_a) by adding a penalty rate h_p where the overall holding cost is given by $h = h_a + h_p$. Similarly the back-order is closely related to availability. The back-order cost (from the customer's viewpoint) is given as $b = b_a + b_p$ where b_a is the actual back-order cost to the customer and the customer has to pick the back-order penalty rate (b_p) as part of the incentive for the contractor in the contract design. The minimization problem that must be solved can be described by,

$$\underset{b_p, h_p}{\text{Min}}\ \text{E}\left(\sum_{t=0}^{T} \varphi[x(t), u(t)] + g(T)\right) \tag{19.3}$$

such that

$$\begin{aligned} &\varphi[x(t), u(t)] = Max(hx(t), -bx(t)) + s|u(t)| \\ &x(t+1) = x(t) + u(t) - d(t) \end{aligned}$$

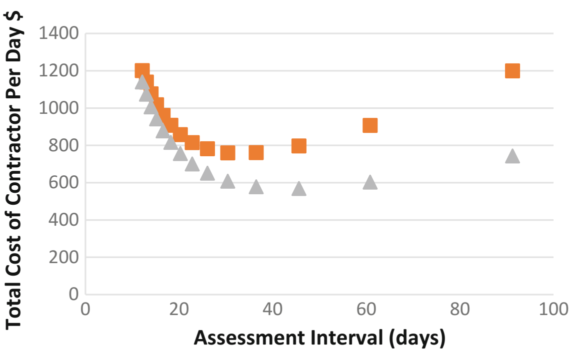
In which $x(t) \in R$ is the quantity of spare parts available in the inventory at time t, for $t = 0$ to T, with $x(t) < 0$ represents a backlog of $|x(t)|$ spare parts. The demand for this component's spare parts at time t is denoted by $d(t)$, which is the stochastic demand for the inventory parts (derived from the failure distribution of the part). The quantity of the spare parts shipped to the inventory at time t is denoted by $u(t)$ with $u(t) < 0$ representing the amount of $|u(t)|$ is shipped back to the factory from the inventory (this particular situation is not allowed in our case). $\varphi[x(t), u(t)]$ is the contractor's cost to operate the inventory given the demand $d(t)$ and order-size u(t). The contractor tries to minimize this cost of operating the service inventory and towards achieving this goal with minimum costs. We also assume that the contractor's shipping costs are fixed and cannot be changed significantly (i.e., there is no uncertainty in it). There is also cost associated with each assessment independent of the inventory size that we reflect in $g(T)$. In this case study, the assessment cost is the cost of testing each torpedo to determine the necessity for repair or replacement.

Assuming the following inputs: $s = 6$($/spare), $g = 12$($/assessment), $6 < h < 24$($/spare/day), $6 < b < 24$($/spare/day) and $5 < d(t) < 1000$ number of failures per day, solving for the optimum values of the two contractual parameters of back-order penalty and inventory level incentives, we need to search the entire feasible design space.

To search the design space we use an optimal affine controller developed in [35] to model the optimal decision making of the contractor. This model is well suited to address the correlated-in-time demand distribution to address a general assumption about the common failures in a fleet of systems. The result of this optimization is shown in Fig. 19.7. For example for the case of intervals of 30 days with 12 assessment a year, the expected cost to the contractor is about $600 per day to support the inventory for the variable number of failures when the assessment cost is low ($12/assessment). It should be noted that this result is based on the unit of time being 1 day, but the result scales for any unit of time.

In Fig. 19.7, the performance of the contractor over a wide range of parameters and demand is shown. There is an optimal customer assessment interval at which, despite all the uncertainties in failure rate and penalty costs, the cost to the contractor to maintain the inventory availability will be minimized. Also, it can be observed that the contracts in which the contractor performance is being assessed based on a longer time interval (every 60 days) has almost the same performance as when the assessment runs every 20 days when the assessment cost is large. Without any assessment cost there exists a diminishing advantage of performing the assessment on a more frequent basis. The impact of uncertainty in information in different stages

Fig. 19.7 Cost per year using various assessment intervals: *squares-assessment* cost high ($20/assessment) and *triangles assessment* cost low ($12/assessment)



of life-cycle can be included within the existing model and a different conclusion could be made for different phases of a contract or life-cycle of a PSS.

One can understand that contractor prefers to undergo a larger assessment interval because: (1) more information is more helpful for demand forecasting in each interval, (2) the effect of one period with lower performance on the overall performance will be minor (total performance measurement is more tolerant towards demand variability), and (3) there is more time and opportunity to compensate a sudden change in demand or a shortcoming in performance.

19.5 Discussion

PSS does not, by itself, ensure that the availability of a system will be guaranteed without designing appropriate contractual requirements. Kashani Pour et al. [31] noted that availability-based contract design has remained mostly unexplored for several reasons: (1) availability-based contracts are relatively new, and cost modeling for this class of contract requires new approach [5]; (2) there is not enough publicly available data to help empirical researchers explore their effectiveness [36]; (3) the engineering design process conventionally does not directly model or utilize the contractual and economic aspects of systems including contractor's incentives and level of effort; and (4) usually sustainment activity problems are divided into sub-problems and solved separately (e.g., supply chain, logistics, and maintenance), which does not utilize the potential advantages of an integrated point of view [37].

In this chapter we explored and demonstrated the core essential elements for the development of informed TES strategies for two classes of PSS that are managed via availability-based contracts. With a systematic and model-based contract design process, not only the pricing and negotiation of the contract requirements can be informed based on the total life-cycle requirements of the PSS; but also, this process can be used to validate the value-added of a variety of technologies including condition monitoring (i.e., CBM and PHM), and obsolescence management.

Acknowledgements Funding for the wind turbine real options work in this chapter was provided by Exelon for the advancement of Maryland's offshore wind energy and jointly administered by MEA and MHEC, as part of "Maryland Offshore Wind Farm Integrated Research (MOWFIR): Wind Resources, Turbine Aeromechanics, Prognostics and Sustainability". Additional funding for the study of contract engineering was provided by the Naval Postgraduate School (Grant Number N00244-16-1-0003) and the National Science Foundation Division of Civil, Mechanical and Manufacturing Innovation (Grant No. CMMI1538674).

References

1. Gansler JS, Lucyshyn W, Vorhis C (2011) Performance-based services acquisition. UMD-CM-11-005, University of Maryland School of Public Policy, February, p 83

2. Levitt T (1972) Production-line approach to service. Harv Bus Rev 50(5):41–52
3. Meier H, Roy R, Seliger G (2010) Industrial Product-Service Systems—IPS^2. CIRP Ann Manuf Technol 59(2):607–627
4. Sundin E (2009) Life-cycle perspectives of product/service-systems: in design theory. In: Sakao T, Lidnahl M (eds) Introduction to product/service-system design. pp 31–49
5. Datta PP, Roy R (2010) Cost modelling techniques for availability type service support contracts: a literature review and empirical study. In: Proceedings of 1st CIRP Industrial Product-Service Systems (IPS^2) conference. Cranfield University, p 216
6. Arora V, Chan FTS, Tiwari MK (2010) An integrated approach for logistic and vendor managed inventory in supply chain. Expert Syst Appl 37(1):39–44
7. Ferguson J, Sodhi M (2011) Using simulation for setting terms of performance based contracts. In: Hu B, Morasch K, Pickl S, Siegle M (eds) Operations research proceedings 2010. Springer, Berlin, p 465
8. Bowman RA, Schmee J (2001) Pricing and managing a maintenance contract for a fleet of aircraft engines. Simulation 76(2):69–77
9. Kim S, Cohen MA, Netessine S (2007) Performance contracting in after-sales service supply chains. Manag Sci 53(12):1843–1858
10. Gabriel SA, Sahakij P, Balakrishnan S (2004) Optimal retailer load estimates using stochastic dynamic programming. J Energ Eng 130(1):1–17
11. Murthy DNP, Asgharizadeh E (1999) Optimal decision making in a maintenance service operation. Eur J Oper Res 116(2):259–273
12. Haddad G, Sandborn PA, Pecht MG (2014) Using maintenance options to maximize the benefits of prognostics for wind farms. Wind Energy 17(5):775–791
13. Lei X, Sandborn PA (2016) PHM-based wind turbine maintenance optimization using real options. Int J Progn Health Manag 7(1)
14. Vestas (2013) 3 MW platform. http://pdf.directindustry.com/pdf/vestas/3-mwplatform/20680-398713.html. Accessed 6 June 2016
15. Bruck M, Goudarzi N, Sandborn P (2016) A levelized cost of energy (LCOE) model for wind farms that includes power purchase agreement (PPA) energy delivery limits. In: Proceedings of the ASME power conference, Charlotte, NC, USA
16. Lei X, Sandborn P, Bakhshi R, Kashani-Pour A, Goudarzi N (2015) PHM based predictive maintenance optimization for offshore wind farms. In: Proceedings IEEE conference on prognostics and health management (PHM), Austin, TX, June 2015
17. Nowicki D, Kumar UD, Steudel HJ, Verma D (2006) Spares provisioning under performance-based logistics contract: profit-centric approach. J Oper Res Soc 59(3):342–352
18. Jazouli T, Sandborn P, Kashani-Pour A (2014) A direct method for determining design and support parameters to meet an availability requirement. Int J Perform Eng 10(2):211–225
19. Jazouli T, Sandborn P, Kashani-Pour A (2014) A direct method for determining design and support parameters to meet an availability requirement—parameters affecting both downtime and uptime. Int J Perform Eng 10(6):649–652
20. Jin T, Wang P (2011) Planning performance based logistics considering reliability and usage uncertainty. Working Paper for Ingram School of Engineering, Texas State University, San Marcos, TX
21. Sharma DK, Cui Q, Chen L, Lindly JK (2010) Balancing private and public interests in public-private partnership contracts through optimization of equity capital structure. Transp Res Rec: J Transp Res Board 2151(1):60–66
22. Hamidi M, Liao H, Szidarovszky F (2014) A game-theoretic model for outsourcing maintenance services. Proceedings—annual reliability and maintainability symposium. doi:10.1109/RAMS.2014.6798514
23. Deng Q, Zhang L, Cui Q, Jiang X (2014) A simulation-based decision model for designing contract period in building energy performance contracting. Build Environ 71:71–80
24. Hong S, Wernz C, Stillinger JD (2015) Optimizing maintenance service contracts through mechanism design theory. Appl Math Model. doi:10.1016/j.apm.2015.07.009

25. Zhu Q, Fung RYK (2012) Design and analysis of optimal incentive contracts between fourth-party and third-party logistics providers. In: Proceedings of the automation and logistics (ICAL). IEEE, pp 84–89. doi:10.1109/ICAL.2012.6308175
26. Arora V, Chan FTS, Tiwari MK (2010) An integrated approach for logistic and vendor managed inventory in supply chain. Expert Syst Appl 37(1):39–44
27. Wang W (2010) A model for maintenance service contract design, negotiation and optimization. Eur J Oper Res 201(1):239–246
28. Zhao H, Yin Y (2011) A dynamic mechanism of risk allocation of construction project from perspective of incomplete contract theory: a theoretical model. Ind Eng Eng Manag 1816–1820
29. Rodrigues D, Erkoyuncu J, Starr A, Wilding S, Dibble A, Laity M (2015) Review of the modelling approaches for availability contracts in the military context. Procedia CIRP 30:451–456
30. Hockley CJ, Smith JC, Lacey LJ (2011) Contracting for availability and capability in the defence environment. In: Ng I, Parry G, Wild P, McFarlane D, Tasker P (eds) Complex engineering service systems: concepts and research. Springer, New York, pp 237–256
31. Kashani Pour AR, Sandborn P, Cui Q (2016) Review of quantitative methods for designing availability-based contracts. J Cost Anal Parametr 9:69–91
32. Sols A, Nowicki D, Verma D (2008) n-dimensional effectiveness metric-compensating reward scheme in performance-based logistics contracts. Syst Eng 14(3):305–326
33. Erkoyuncu JA, Roy R, Shehab E, Wardle P (2009) Uncertainty challenges in service cost estimation for product-service systems in the aerospace and defence industries. In: Proceedings of the 1st CIRP Industrial Product-Service Systems (IPS^2) conference, April 2009
34. Knight JT, Singer DJ (2014) Applying real options analysis to naval ship design. In: Proceedings of ASNE Day
35. Skaf J, Boyd SP (2010) Design of affine controllers via convex optimization. IEEE Trans Autom Control 55(11):2476–2487
36. U.S. Government Accountability Office (2008) Improved analysis and cost data needed to evaluate the cost-effectiveness of performance-based logistics
37. Shen ZJ (2007) Integrated stochastic supply-chain design models. Comput Sci Eng 9 (2):50–59

Part VII
Autonomous Maintenance and Product Support

Chapter 20
Application of Open Source Hardware to the Development of Autonomous Maintenance Support Systems

Michael Knowles, David Baglee and Pankaj Sharma

Abstract Autonomous maintenance systems offer organizations the opportunity to embed state of the art maintenance tools and techniques with minimal operator input supported by a range of technologies such as condition monitoring sensors and techniques, intelligent data processing systems and smart prognostic algorithms. Many companies perceive that autonomous maintenance is difficult to achieve due to a lack of understanding of the infrastructure required to support such an approach as this requires an understanding of the fundamental principles of electronic instrumentation, processing and communication techniques, alongside the ability to create and integrate the appropriate software and firmware. Open source hardware has received attention in recent years as it allows a range of users to create sophisticated applications quickly using readily available components and modules. Such platforms are supported by a range of library software designed to further accelerate and simplify the development process. These products have attracted much attention from hobbyists but are now attracting attention in their own right from potential industrial users. However the reliability of these systems in an industrial environment remains a concern. In this chapter the benefits of applying open source technologies to create an autonomous maintenance system will be examined alongside the perceived and actual barriers limiting their uptake. The required enablers to achieve the potential benefits will then be explored leading to a detailed roadmap identifying what needs to be achieved for the significant industrial potential of these devices and systems to be realized.

M. Knowles (✉)
School of Engineering, University of Sunderland, Sunderland, UK
e-mail: michael.knowles@sunderland.ac.uk

D. Baglee
Institute for Automotive and Manufacturing Advanced Practice,
University of Sunderland, Sunderland, UK
e-mail: david.baglee@sunderland.ac.uk

P. Sharma
Indian Institute of Technology Delhi, New Delhi, India
e-mail: pankajtq@gmail.com

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_20

20.1 Introduction

Maintenance strategies in industry have evolved as the underlying technologies have developed. Prior to the 1950s, *reactive maintenance* was prevalent which was subsequently replaced by *preventive maintenance*. During the 1960s, *productive maintenance* became more popular. It was, however, in the 1970s that *Total Productive Maintenance (TPM)* emerged as the preferred form of maintenance. Seiichi Nakajima, vice-chairman of the Japanese Institute of Plant Engineers (JIPE), is known as the founder of TPM. TPM is designed to maximize equipment effectiveness and thus improve overall efficiency by establishing a comprehensive productive-maintenance system covering the entire life of the equipment, spanning all equipment-related fields (planning, use, maintenance, etc.) and, with the participation of all employees from top management down to shop-floor workers, to promote productive maintenance through motivation management or voluntary small-group activities [1].

One of the main arms of TPM is Autonomous Maintenance (AM), known as *Jitshu Hozen* in Japanese. The success of TPM is largely dependent on the success of AM. Autonomous Maintenance is a practical application of TPM which aims to promote a culture in which operators feel that they "own" their machines, learn much more about them, and in the process release skilled trades to concentrate on problem diagnosis and equipment improvement projects [2]. AM ascertains the roles and tasks of production operators so they can perform easy daily maintenance activities alongside yet distinct from planned maintenance. In other words, AM is designed to oblige production operators to maintain their own equipment independently without notice or instruction from the maintenance department [3].

20.1.1 The Role of Computers in Autonomous Maintenance

The advent of computers into asset maintenance has brought about major changes in the field in the last two decades AM, as defined by Seiichi Nakajima in the 1970s, placed the emphasis on cleaning, inspection and minor repairs/adjustments by the operator [4]. A part of this AM system was diagnosis and prognosis of the fault using predictive maintenance tools. Initially, the most widely available tool with the operator was his experience on the machine. The practice was to focus more on the social aspect of the issue by motivating the operator to feel ownership of the machine and become familiar with it so that any impending failure could be "sensed" by the operator. With the increase in computing powers and decrease of the cost of data acquisition and analysis, it has become possible to equip the operator with modern tools of condition monitoring sensors and techniques, intelligent data processing systems and smart prognostic algorithms. Condition-based autonomous maintenance methods are being widely used in industry today. Some of these are manually controlled and provide information to the operators through cameras and sensors.

Others are semi-autonomous/fully autonomous in their monitoring process. The more modern maintenance systems are capable of self-monitoring/-healing/-repair.

20.1.2 Low-Cost Embedded Computing Systems

Recent years have seen the introduction of a range of microprocessor platforms and single board computers which are designed to encourage those from non-traditional electronic engineering backgrounds to become familiar with this technology [5, 6]. Such systems are supported by a range of tools to simplify and accelerate the software development process. The simplicity of programming coupled with the low cost of these platforms has made them attractive to a range of users including hobbyists, students and educational establishments. The popularity of such hardware and the evident potential it offers means that industrial users are now seeing a range of potential applications.

The term ‘Open Source Hardware’ refers to circuits and assemblies whose design can be freely modified and used in any setting. The definition proposed by the Open Source foundation is “hardware whose design is made publicly available so that anyone can study, modify, distribute, make and sell the design or hardware based on that design” [7]. The open source hardware movement is supported by a range of licenses that can be applied to the reuse of designs and associated documentation and software etc. It should be noted, however, that most licenses are based around the concept of ‘copyright’ and thus do not extend to the hardware itself or any aspect of its manufacturing [7].

A number of single-board microprocessor platforms are supported by additional modules which can be connected to the processor offering a range of functionalities such as motor control, LCD displays including touch screens and various environmental sensors. Examples of such ranges include Arduino ‘Shields’, which share a common footprint and pin layout allowing easy physical integration with the main processor board for a number of different accessories. These entire devices offer a degree of ‘plug and play’ in their application minimizing the need to develop bespoke printed circuit boards (PCBs) and other interconnecting hardware. The ‘open source’ nature of the Arduino platform in particular has fostered a culture of innovation where a worldwide community of hardware developers have created a diverse range of low cost hardware add-ons including many which are manufactured at low cost in the Far East and marketed worldwide via the World Wide Web (WWW). The quality and extent of the documentation for low cost hardware modules is, however, variable.

The Arduino platform was launched in 2005 and offered a single board platform based initially on the Atmel AVR microcontroller. One of the principal target markets for the Arduino platform was hobbyists—a community in which the platform has been extremely popular. In addition to the official product which is affordable to all, the Arduino hardware is open source in nature meaning low cost ‘clones’ are also available which has contributed to the popularity of the platform. Since launch, a number of variants have been developed using a range of different processors and with a range of different capabilities.

The primary advantage of the Arduino platform is the ability it offers to users to develop applications without any hardware development, without the need for additional programming hardware. This removes a significant barrier to those wishing to use microprocessor systems but who do not have the time, infrastructure or knowledge to develop the underlying hardware [5].

Following the success of Arduino, a number of similar single-board systems have appeared on the market. The BeagleBoard was launched by Texas Instruments in 2008 and can function as a single-board computer. The Raspberry Pi platform was first launched in 2012 and also functions as a single-board computer. More recently the Arduino movement has released new versions such as the Arduino Yún with similar capabilities.

20.1.3 Embedded Autonomous Maintenance Systems

Low cost embedded systems such as those described above offer a significant opportunity for the development of systems and devices which can be used to support localized, operator led maintenance activities. In this chapter the nature of AM will be investigated. Enabling technologies and concepts will then be explored such as the functionality of the available hardware, its reliability and robustness for the proposed application and the impact of current research areas such as the “Internet of Things” and “Big Data”.

20.2 Literature Review

20.2.1 Examples of Autonomous Maintenance

Autonomous Maintenance (AM) is one of the pillars of Total Productive Maintenance (TPM). TPM is a partnership between maintenance and production functions in the organization to improve product quality, reduce waste, reduce manufacturing cost, increase equipment availability, and improve the company’s state of maintenance [8]. There is a plenty of work available in literature that deals with benefits of a successful implementation of TPM in the industry [9–14]. Similarly, there is ample research on the barriers to implement a successful TPM program [15, 16]. Ahuja and Khamba [17] summarizes the problems in TPM implementation as cultural resistance to change, partial implementation of TPM, overly optimistic expectations, lack of a well-defined routine for attaining the objectives of implementation (equipment effectiveness), lack of training and education, lack of organizational communication, and implementation of TPM to conform to societal norms rather than for its instrumentality to achieve world class manufacturing [18, 19].

AM has come a long way from what it was described by Nakajima [4]. AM of the modern times can be categorized into two main categories. The first is the "self-monitoring and repair/healing" category. The second focuses upon automating maintenance practices within organizations, with a particular focus upon the use of autonomous robotics to aid, guide or take over current maintenance tasks undertaken by human engineers/workers [20]. The next two paragraphs deal with the literature in these two categories.

Literature in the self-* category deals with the ability of the machines to self-test, self-monitor, self-diagnose, self-heal and self-repair. Important steps are; firstly, to make the machine aware of its current status; secondly, to know how it is expected to behave and thirdly, to arrive at a proper diagnosis by sensing the difference between the two [21–23]. Bell et al. [24] describe three methods to do automated diagnosis; Model based, Bayesian belief network based and case based reasoning. The corrective action in the machine can then take place either through self-healing or self-repair. Self-healing involves the part to heal the damage from inside. This can be done with the help of self-healing materials, software or mechanisms. A detailed survey on this can be read in [25]. Self-repair; on the other hand, means that the system has the ability to partially or fully fix a given fault. This can be done through self-reconfiguration [26]. A single spare module can be designed that is able to replace a host of primary modules in spite of having different structure. Alternatively this concept of self-repair through self- reconfiguration does not necessarily require additional materials; instead performance can be sacrificed to ensure continued functionality utilizing only the currently available resources [24].

Most prominent applications of the second category have been found in maintenance of railways, pipelines, sewerage systems. Turner et al. [27] describe a project in novel sensing, scheduling, and decision-making strategies customized for the automated planning of maintenance activities within the rail industry. Three key areas of research in the project are sensor fusion, planning & scheduling and costing. Dadashi et al. [28], Bye [29] and Dadashi et al. [30] have studied the applications of handheld computers in the rail industry. Robot agent based technologies are considered as an attractive alternative for fully/semi-autonomous pipeline monitoring and inspection. Moreover, robot agent-based technologies free the engineers from the confinement of pipeline inaccessibility, environment hazardousness, and system scalability [31]. In the field of autonomous maintenance of pipelines, Wang et al. [32] developed a novel autonomous in-pipe robot to perform the preventive point repair for long-distance offshore oil pipelines. The autonomous in-pipe robot performs online ultrasonic inspection for pipe wall thickness, and the original inspection data are stored in large capacity hard disk. Through the offline data analysis by the data analysts and the software tool, the pipeline health status is known. If server defects lie there, the in-pipe robot is introduced into the pipeline once more to indicate the defect's location to the maintenance ship. Kim et al. [31] propose a novel Radio Frequency Identification (RFID)-based Autonomous maintenance system for

Pipelines which combines robotic, sensing, and RFID technologies for efficient and accurate inspection, corrective repair, and precise geo-location information. Other works in this field include [33, 34]. Kirkham et al. [35] and Hertzberg and Kirchner [36] describe semi-autonomous inspection robots for sewer pipes. Robots have been of great help in carrying out maintenance where climbing vertical structures is involved. Schmidt and Berns [37] conduct a comprehensive literature review on climbing robots for inspection and maintenance of vertical structures.

20.2.2 Role of Autonomous Maintenance Within TPM

AM is a major constituent of the TPM approach. AM simply implies that the inspection be incorporated with production when possible, thus allowing people to inspect their own work and learn from their mistakes. This also reduces the number of communication links across departmental boundaries. The fewer the variances that are imported from the place where they arise, the fewer the levels of supervision and control that are required [38]. AM is one of the eight pillars of TPM that aims at developing operator ownership. The operator performs day-to-day tasks to be able to develop skills and in turn mastery of the equipment [39]. As operators are trained, they begin to inspect and maintain the equipment and perform basic maintenance tasks. This allocation of maintenance tasks to production operators frees up, provides more time or something similar up time for maintenance personnel to perform long-term improvement efforts and plan maintenance interventions [40]. Nakajima [4] presented a framework for a four phase implementation of an AM and Planned Maintenance approach. Mckone and Weiss [40] later included a fifth phase in the framework. These five stages and the steps taken as part of the AM approach are enumerated in Table 20.1.

The first phase of the plan deals with attaining control over the machine by eliminating the sources of problem. In the second phase, operators are trained to carry out general inspection of the machine and follow the standards set in phase 1. Phase 3 is related to carrying out autonomous inspection based on preventive

Table 20.1 AM in the five phases of TPM development [40]

	Phase 1 Reduce life span variability	Phase 2 Lengthen average life span	Phase 3 Estimate life span	Phase 4 Predict life span	Phase 5 Design life span
Autonomous maintenance steps	1: Basic cleaning 2: Eliminate source of problem 3: Set standards	4: General inspection of equipment	5: Autonomous inspection	6: Maintenance of quality 7: Autonomous maintenance	8: Process improvement and design team 9: Implement in all support areas

maintenance schedules. In phase 4, predictive maintenance tools such as condition monitoring are available that help carry out high quality maintenance. In phase 5, design teams made up of engineers, maintenance workers, and operators prepare equipment so that cleaning and inspection standards are established and personnel are trained to produce effectively upon roll-out. Phase 5 decisions also consider other non-maintenance systems, such as spare parts, raw materials, and production scheduling that impact the equipment productivity and quality [40].

There have been major advancements in the field of autonomous monitoring and maintenance of the machines. Sensor networks have witnessed a rapid growth due to the development of inexpensive sensing devices and communication technologies and are used for several applications such as agriculture, military, health care, and pipeline monitoring [31]. The use of robotics has also increased in the recent years. The application of robots can be restricted to being non-autonomous being guided by the humans for navigation to the fault and conduct of inspection through the robot by the operator [41]. These non-autonomous robots (e.g., remote-controlled) are equipped with cameras, sensors, and master-slave controlled manipulators [42]. The robots are termed as semi-autonomous where the navigation part is carried out autonomously but the decision making and repairing is done by the humans [43]. The autonomous robots have the capability to carry out navigation, inspection and repair.

AM implied that the operator is skilled and trained to autonomously undertake inspection and repair of his machine. AM has evolved to mean that the machine itself is able to monitor and repair faults in it. Researchers are looking for methods of self-monitoring, self-testing and self-diagnosis of faults by the machines.

20.2.3 Barriers to Autonomous Maintenance

While the benefits of Autonomous Maintenance are clear, there are certain barriers in implementing a truly AM system capable of carrying out the required actions. These barriers are listed below.

- **Non-modularity of the design of the assets**: AM demands high level of modularity in the assets. It helps in diagnosis, fault isolation and disassembly. The damaged or in need of service components can be maintained or repaired after disassembly at a later date at separate and more ideal facilities. Incorporating such an ideal of modularization into design could be an alternative route to building ever more complex autonomous robots [20]. Modularity will also help in incorporating redundancy of parts by having spare modules that are of same type as the degenerate module; or have single spare module that has the capability to replace any defective module [24].

- **Embedding of sensors in the existing assets**: Majority of the AM tasks involve some level of monitoring by the sensors. In case an AM plan has to be implemented in existing assets, it is a challenge to embed these sensors without adversely impacting the asset itself. It implies that in certain cases, there may be a need to re-design the asset itself. Other option would be to invest more on robotic monitoring systems that are external to the asset.
- **Adapting to non-deterministic environment**: Maintenance is a non-deterministic activity. The situation and environment of the maintenance activity is likely to change; and at an unknown rate. These varying scenarios present a vast number of possible decisions for the AM system to make. The AM system needs to have cognitive ability and intelligence to pick the most optimal decision.
- **Diagnosis**: Diagnosis of the fault has traditionally been the most difficult step in maintenance. The same is true even in the case of AM systems. The machines are getting complex and hence there are a vast number of possible system states. There are further challenges of not having adequate confidence in the diagnosis. An additional step is required in which the diagnosis must be confirmed, to avoid undesirable events such as 'good' components being unnecessarily removed or routed around [24].
- **Financial barrier**: Any new technology will find practical application in the industry if there are cost benefits of doing it. The investments in modern AM systems are large; larger than the possible returns of reduced maintenance and downtimes. Due to this reason, the application is limited to most critical systems and places where it is hazardous for the human beings to operate. This is the possible reason why the research is focussed more on AM of vertical structures and underwater pipelines. Low cost systems discussed in this chapter attempt at breaking this barrier and making AM more cost effective.
- **Social barriers**: All organizations are sociotechnical systems. Organizational objectives are best met not by the optimization of the technical system and the adaptation of social system into it, but by the joint optimization of the technical and the social aspects [38]. Sociotechnical theory has at its core the notion that the design and performance of new systems can be improved, and indeed can only work satisfactorily, if the 'social' and the 'technical' are brought together and treated as interdependent aspects of a work system [44]. AM may have evolved technically over the years, but the social implications in successfully implementing it in an industry have not changed. The workers may see it as a means to cut down jobs. There are extra skills required by the operators to operate semi-autonomous robots. This in turn implies greater investments in training. The top management may be committed to implement such a program, but it must have the patience to implement it in stages. The success and the lessons learnt in a stage should be utilised to fuel the subsequent stages of implementation.
- **Need of a multi-disciplinary workforce**: There is a need to have a multi-disciplinary workforce; particularly in the decision making echelon that has the knowledge of engineering, computing, analytics, automation, design and

production. This may be difficult in certain industries. It also results in increasing the costs of hiring and retaining a potentially more qualified workforce.

- **High risk of disruptions**: There is a high-risk associated with new technology and system implementation. The implementation of an AM system "may negatively impact operations while personnel are achieving competency or fail to operate as originally intended" [45].

20.3 Role of Low Cost Embedded Systems in Facilitating Autonomous Maintenance

Low cost embedded microprocessor based platforms offer the potential to develop solutions which address many of these barriers. Low cost platforms and the rapid system development they offer have the potential to de-risk the development of localized condition monitoring systems which can be readily adapted to operate on a range of assets. This means that condition monitoring systems which adapt to non-modularity can easily be developed to suit particular 'bespoke' equipment and systems. Furthermore sensor systems can be designed more flexibly to address impacts on the normal operation of the asset.

The computational power of platforms such as Raspberry Pi, Arduino and BeagleBoard mean that a considerable level of intelligence can be embedded at the level of a particular asset. Historically, condition monitoring systems often relied on centralized processing, rule setting and data storage to handle the data from many assets thus decoupling the operator from the process and undermining the principles of AM. This can now be avoided through the use of localized processing and storage. Furthermore, the connectivity offered by such platforms means that the benefits of interconnection are not lost meaning that centralized monitoring can still take place alongside operator led processes.

The accessibility of such platforms means that condition monitoring can become truly integrated within the AM and TPM methodologies rather than being something practiced by specialist technicians. Any machine operator will have the ability to understand, develop and enhance the condition monitoring platform with minimal training if the appropriate development tools and aids are in place.

20.3.1 Integration Between Localized and Centralized Systems

Big Data has been identified by the scholars as "the next big thing in innovation" [46] and the "new paradigm of knowledge assets" [47]. Wu et al. [48] describes Big Data characteristics through HACE theorem; large-volume and Heterogeneous, Autonomous source with distributed and decentralized control, and seeks to explore Complex and Evolving relationships among data. High dimensionality of this big data helps in predicting the future by establishing relationships between variables. Autonomous data sources with distributed and decentralized controls are a main characteristic of Big Data applications. Being autonomous, each data source is able to generate and collect information without involving (or relying on) any centralized control [48]. The real world relationships amongst the variables are complex and evolving. Large sample size of the data helps in understanding this complex relationship by firstly, exploring the hidden structures of each subpopulation of the data, which is traditionally not feasible and might even be treated as 'outliers' when the sample size is small; and secondly, by extracting important common features across many subpopulations even when there are large individual variations [49]. Big Data Analytics (BDA) has helped organizations and systems to analyze large volume of high dimension data from a large sample size collected by autonomous decentralized sensors. BDA has the capability to predict the future course of events by establishing trends in otherwise unrelated variables.

AM involves collection of data by autonomous sensors placed on the machines. There are other sources of data which reside in the ERP systems, process control systems, smart devices etc. The employees and customers are adding to this ever increasing data deluge through social media. The data need not always be numerical data. Keywords being used in the reviews, tweets, logbooks, inspection sheets, job cards and other such sources can be picked up to derive meanings. This data needs to be analysed along with data from several such sources distributed across the globe to discover hidden relationships that can help in predicting the future events like equipment failure. Big data and predictive analytics can also help by providing operation warnings and automatic interventions. Predictive analytics software use fast algorithms scalable to massive data with high dimensionality to detect subtle variances for each piece of equipment, which often warn of impending problems that might have gone unnoticed otherwise.

The machines have traditionally been fitted with sensors that transmit data through proprietary communication protocols to proprietary control interfaces that are generally located within the premises of the organization. With emergence of Internet of Things (IoT), these sensors can now communicate using the standard internet protocols, using open standards, making the data available to a wide range of users on the internet. The cost of computing as well as the cost of the sensors is decreasing rapidly. This has increased the potential to embed the machines with more sophisticated sensors and processors. As these sensors are now connected through standard internet protocols, the volume of data has increased that can lead to real-time analysis.

20.3.2 *Enablers Required for Realization of Low Cost Autonomous Maintenance Support*

In order for Low Cost Hardware Platforms to be effectively and optimally applied to the support of AM, the following barriers and issues must be addressed:

- **Reliability and Resilience**. One of the major issues affecting low cost "off the shelf" hardware is the fact the design of these devices is often heavily focused on non-industrial applications. A number of reliability related issues exist for these platforms and devices:
    - The reduction in size of the devices involved means that there is a greater risk of transient faults affecting the reliability of the system [50].
    - Such devices are often developed with a minimal level of protection against various stresses such thermal conditions, electrical transients and electrostatic discharge (ESD). While there are well-proven techniques to mitigate against these issues in high risk situations such as the use of protection diodes, shielding, protective grounding and transient protection devices [51], these are often overlooked in low cost applications.
    - Electronic systems are often incorrectly assumed to be immune to aging process due to the absence of moving parts. While wear rates are potentially lower, several studies, e.g. Knight et al. [52] have indicated that faults can occur—even in low power systems such as electronic systems in automotive environments. Thus systems operating in a manufacturing context are likely to be equally susceptible to age related degradation processes.
- **Cost**. A further enabler for the industrial use of off the shelf microprocessor systems for AM applications lies in the cost. While the 'core' platforms themselves are often low-cost, the required additional hardware should be minimized since this is often where costs can accumulate. This can make the selection of platform critical since an incorrect choice e.g. of a platform with no intrinsic communications capabilities can be rapidly undermined by the need to add hardware to achieve wider integration. Other additional hardware which may add cost in condition monitoring applications include signal processing and conditioning systems involved in integrating with sensors and other data sources. A fundamental trade-off thus exists between having a platform that is intrinsically flexible enough to benefit from mass production and one which is not so heavily featured as to drive up cost.
- **Development Support**. One of the key factors in the success of the Arduino platform is the common and easily accessible development environment which is offered and which is based on widely used programming languages such as Java, C and C ++. Such platforms by their very nature do not provide support for application specific functionality since these are likely to be underutilized. While additional software libraries exist to perform many such functions, these are often lacking in the degree of verification required for industrial users to have confidence in them.

20.4 Roadmap for Achieving Low Cost Autonomous Maintenance Support

If autonomous maintenance is to be successfully supported by off the shelf hardware with the minimum level of additional development, the following research outcomes need to be achieved:

- **Identification of core functionality to ensure efficient production and minimum additional hardware**. Ensuring the correct balance is struck between versatility and simplicity of a platform will be a critical consideration in the design of an open-source computing platform that is suitable for supporting AM and localized condition monitoring. A key issue will be the choice of microprocessor or microcontroller. Thus analyzing the needs implied by the potential applications will be critical in this process.
- **Identification of key reliability issues in electronic systems/industrial environments**. Once the core functionality and key devices have been selected a robust yet affordable platform will need to be generated taking full advantage of established approaches to designing robust electronic systems. This will require research into the environmental conditions that might be reasonably expected in the desired application, as well as detailed prototyping, testing and evaluation to demonstrate the reliability of the hardware.
- **Identification of core software development support systems**. This will require consideration of the existing skills and knowledge of the operators and an understanding of how they may wish to configure and use such systems to support their role in autonomous maintenance. Systems and approaches to support rapid development and reconfiguration are thus required while maintaining the ability for advanced users to fully exploit the capabilities of the platform.
- **Identification of core technical knowledge and skills required for machinery operators to actively engage in AM**. It is likely to that some upskilling will be required but this should focus on how the platform can support AM rather than on the technical details of the platform. Thus abstracting the details of the implementation and developing a platform that encourages enhancement of operator skills is a key objective.

20.5 Conclusions

Reliability and efficiency of equipment is a key factor in productivity for many manufacturing and engineering operations. While approaches such as TPM are well established, their role within an asset management approach continues to evolve in the light of continuous developments in supporting technologies.

Off the shelf embedded platforms offer considerable potential for further evolving the role of operators in maintaining their own assets. In order to fully maximize this potential the technology needs to be carefully developed to ensure that it is robust, reliable, easy to develop and apply and of commercial benefit. Furthermore, and perhaps more importantly, it must be seen and perceived to possess these attributes by those at all levels within organizations with responsibility of maintenance, from senior management down to operators, if widespread uptake is to be achieved.

References

1. Tsuchiya S (1992) Quality maintenance: zero defects through equipment management. Productivity Press, Cambridge, MA
2. Willmott P (1994) Total quality with teeth. TQM Mag 6(4):48–50
3. Min CS, Ahmad R, Kamaruddin S, Azid IA (2011) Development of autonomous maintenance implementation framework for semiconductor industries. Int J Ind Syst Eng 9(3):268–297
4. Nakajima S (1988) TPM development program. Productivity Press, Cambridge, UK
5. Severance C (2014) Massimo Banzi: building Arduino. Computer 47(1):11–12. doi:10.1109/MC.2014.19
6. Hill L, Steven C (2013) Using a low cost open source hardware development platform in teaching young students programming skills. In: Proceedings of the 14th annual ACM SIGITE conference on information technology education, pp 63–68. doi:10.1145/2512276.2512289
7. English A (2015) Open Source Hardware—an iGEM Guide. University of Cambridge. http://2015.igem.org/wiki/images/d/db/CamJIC-Practices-Guide.pdf
8. Rhyne DM (1990) Total plant performance advantages through total productive maintenance. Conference proceedings APICS, Birmingham, pp 683–686
9. Windle WM (1993) TPM: more alphabet soup or a useful plant improvement concept? Plant Eng 47(2):62–63
10. Chowdhury C (1995) NITIE and HINDALCO give a new dimension to TPM. Udyog Pragati 22(1):5–11
11. Ireland F, Dale BG (2001) A study of total productive maintenance implementation. J Qual Maint Eng 7(3):183–191
12. Chan FTS, Lau HCW, Ip RWL, Chan HK, Kong S (2005) Implementation of total productive maintenance: a case study. Int J Prod Econ 95:71–94
13. Sharma KR, Kumar D, Kumar P (2006) Manufacturing excellence through TPM implementation: a practical analysis. Ind Manag Data Syst 106:256–280
14. Ahuja IPS, Khamba JS (2007) An evaluation of TPM implementation initiatives in an Indian manufacturing enterprise. J Qual Maint Eng 13(4):338–352
15. Attri R, Grover S, Dev N, Kumar D (2013) Analysis of barriers of total productive maintenance (TPM). Int J Syst Assur Eng Manag 4(4):365–377
16. Poduval PS, Pramod VR, Raj VPJ (2013) Barriers in TPM implementation in industries. Int J Sci Technol Res 2(5)
17. Ahuja IPS, Khamba JS (2008) Total productive maintenance: literature review and directions. Int J Qual Reliab Manag 25(7):709–756
18. Crawford KM, Blackstone JH Jr, Cox JF (1988) A study of JIT implementation and operating problems. Int J Prod Res 26(9):1561–1568
19. Becker SW (1993) TQM does work: ten reasons why misguided efforts fail. Manag Rev 82 (5):30–34

20. Farnsworth M, Bell C, Khan S, Tomiyama T (2014) Autonomous maintenance for through-life engineering. Through-Life Eng Serv 395–419
21. Frei R (2010) Self-organisation in evolvable assembly systems. PhD thesis, Department of Electrical Engineering, Faculty of Science and Technology, Universidade Nova de Lisboa, Portugal
22. Visinsky ML, Cavallaro JR, Walker ID (1994) Robotic fault detection and fault tolerance: a survey. Reliab Eng Syst Saf 46(2):139–158
23. Laddaga R (1997) Self-adaptive software. Technical report 98–12, DARPA BAA
24. Bell C, McWilliam R, Purvis A, Tiwari A (2013) Concepts of self-repairing systems. Meas Control 46(6):176–179
25. Frei R, McWilliam R, Derrick B, Purvis A, Tiwari A, Di Marzo SG (2013) Self-healing and self-repairing technologies. Int J Adv Manuf Technol 69:1033–1061
26. Stoy K, Christensen D, Brandt D (2010) Self-reconfigurable robots: an introduction. MIT Press, Cambridge
27. Turner CJ, Tiwari A, Starr A, Durazo-Cardenas I, Blacktop K (2015) An intelligent framework and prototype for autonomous maintenance planning in the rail industry. In: Proceedings of the first international conference on vehicle technology and intelligent control systems, Lisbon, Portugal
28. Dadashi Y, Wilson JR, Sharples S, Clarke T (2008) Applications of handheld computers for the rail industry. In: Bust PD (ed) Proceedings of the international conference on contemporary ergonomics (CE2008), Nottingham, UK. CRC Press
29. Bye R (2013) Designing mobile user experiences: disruptive innovation in railway asset information. In: Dadashi N et al (eds) Rail human factors: supporting reliability, safety and cost reduction, Leiden. CRC Press
30. Dadashi Y, Sharples S, Wilson JR, Clarke T (2012) Investigating presentation of rail specific spatial information on handheld computer screens. Pers Ubiquitous Comput 16(8)
31. Kim JH, Sharma G, Boudriga N, Iyengar SS, Prabakar N (2015) Autonomous pipeline monitoring and maintenance system: a RFID-based approach. EURASIP J Wirel Commun Netw 2015:262
32. Wang Z, Cao Q, Luan N, Zhang L (2010) Development of an autonomous in-pipe robot for offshore pipeline maintenance. Ind Robot: Int J 37(2):177–184
33. Henrique VS, Morooka CK, Guilherme IR, Tiago CF, Mendes JP (2005) Leak detection in petroleum pipelines using a fuzzy system. J Petrol Sci Eng 49(3/4):223–238
34. Feng J, Zhang HG (2005) Diagnosis and localization of pipeline leak based on fuzzy decision-making method. Acta Automatica Sinica 31(3):484–490
35. Kirkham R, Kearney P, Rogers K, Mashford J (2000) PIRAT—a system for quantitative sewer pipe assessment. Int J Robot Res 19(11):1033–1053
36. Hertzberg J, Kirchner F (1996) Landmark-based autonomous navigation in sewerage pipes. In: Proceedings of the first Euromicro workshop on advanced mobile robot, pp 68–73
37. Schmidt D, Berns K (2013) Climbing robots for maintenance and inspections of vertical structures—a survey of design aspects and technologies. Robot Auton Syst 61:1288–1305
38. Cherns A (1976) The principles of sociotechnical designs. Hum Rel 29(8):783–792
39. Parikh Y, Mahamuni P (2015) Total productive maintenance: need and framework. Int J Innov Res Adv Eng 2(2):126–130
40. Mckone KE, Weiss EN (1998) TPM: planned and autonomous maintenance: bridging the gap between practice and research. Prod Oper Manag 7(4):335–351
41. Parker LE, Draper JV (1998) Robotics applications in maintenance and repair. Handbook of industrial robotics, 2nd edn
42. Akrout H, Anson D, Bianchini G, Neveur A, Trinel C, Farnsworth M, Tomiyama T (2013) Maintenance task classification: towards automated robotic maintenance for industry. In: 2nd international through-life engineering services conference, Procedia CIRP, vol 11, pp 367–372

43. Dobie G, Spencer A, Burnham K, Pierce GS, Worden K, Galbraith W, Hayward G (2011) Simulation of ultrasonic lamb wave generation, propagation and detection for a reconfigurable air coupled scanner. Ultrasonics 51:258–269
44. Clegg CW (2000) Sociotechnical principles for system design. Appl Ergon 31:463–477
45. O'Donovan P, Leahy K, Bruton K, O'Sullivan DTJ (2015) An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities. J Big Data 2:25
46. Gobble MM (2013) Big data: the next big thing in innovation. Res Technol Manag 56(1): 64–66
47. Hagstrom M (2012) High-performance analytics fuels innovation and inclusive growth: use big data, hyperconnectivity and speed to intelligence to get true value in the digital economy. J Adv Anal 2:3–4
48. Wu L, Yuan L-Y, You J-H (2015) Survey of large-scale data management systems for big data applications. J Comput Sci Technol 30(1):163–183
49. Fan J, Han F, Liu H (2014) Challenges of big data analysis. Natl Sci Rev 1:293–314
50. Gracia-Morán J, Baraza Calvo JC, Gil Tomás DA, Saiz-Adalid L, Gil P (2014) Effects of intermittent faults on the reliability of a reduced instruction set computing (RISC) microprocessor. IEEE Trans Reliab 63(1):144–153. doi:10.1109/TR.2014.2299711
51. Department of Defense (1998) Electronic reliability design handbook
52. Knight I, Eaton A, Whitehead D (2001) Reliability of electronically controlled systems on vehicles. Transport Research Laboratory (TRL) Report

Chapter 21
Design for Zero-Maintenance

M. Farnsworth, R. McWilliam, S. Khan, C. Bell and A. Tiwari

Abstract This chapter looks at the concept of zero-maintenance, in particular how it relates to design. It begins by defining what constitutes zero-maintenance, presenting current research on the themes of autonomous maintenance and self-healing and repair. A wider context of how zero-maintenance affects through-life engineering services is also discussed with a focus on the no-fault found phenomenon. Case studies are then presented for design strategies in self-healing electronics and no-fault found and the failure of design. Finally, a design for zero-maintenance process is outlined and discussed.

21.1 Introduction

Throughout the life time of products and large assets there is a persistent need for numerous interventions in the form of targeted servicing in order to maintain operation efficiency. Whether predictable by nature or triggered by unexpected events, service support occurs at all stages of product lifetime—from initial conception, design and manufacture to through-life operation and ultimately end-of-life. Integral to many of these stages are processes intended to maintain the function of the product with minimal cost overheads. These processes, whether integrated into the product or else applied in the form of planned or unplanned maintenance, are essential to maintaining sustained performance specifications from the product in question. Conception and planning for maintenance-centric processes begins at the design stage, but is all too often neglected in favour of more typical cost and performance factors. However, new design considerations based around mitigating against costs associated with product failure, maintenance and repair are emerging that need to be brought into the fold of future product design. Such considerations inevitably tie into both the cost and performance of products at some point along their life-cycle chain.

M. Farnsworth (✉) · R. McWilliam · S. Khan · C. Bell · A. Tiwari
Cranfield University, Cranfield, UK
e-mail: m.j.farnsworth@cranfield.ac.uk

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_21

This chapter investigates the philosophy of zero-maintenance i.e., the elimination of maintenance-centric costs, and its relation to through-life engineering services for a range of high-value products and assets. The focus is centered on design, and how information gathered from the through-life services chain can be used to better inform design decisions, so as to ultimately mitigate or remove the need for maintenance whilst continuing to address the traditional goals of reducing operational costs and improving performance.

21.2 Zero-Maintenance

The availability and performance of high-value assets is the key driver of through-life engineering services (TES). Without these two factors manufacturers that undertake such contracts will quickly lose revenue and credibility. Loss in availability or drops in performance are highly correlated to system failure, either as a result of direct damage or through degradation due to environmental factors or operational use. Maintenance, either planned or unplanned is a service undertaken to reduce such system failures or return them to original design specifications. An alternative to the undertaking of maintenance is to design in capabilities for self-healing and repair within the asset itself. These capabilities either provide much needed resilience to damage, thus blunting the impact of any unforeseen event that may have otherwise lead to damage, provide processes for repairing damage autonomously or simply allow damaged functions to be replaced.

Together both standard maintenance and self-healing strategies form a powerful approach to maintaining availability and performance. The philosophy of zero-maintenance encompasses these two strategies to ultimately provide the elimination of maintenance-centric costs, and its relation to through-life engineering services for a range of high-value products and assets. The following sections discuss three main topics surrounding zero-maintenance. The first focuses on the use of autonomy and autonomous systems to perform maintenance and remove human intervention. The second on the strategy of self-healing and repair, and finally a broader look at how a zero-maintenance effects wider through-life engineering services and the phenomena of no-fault-found (NFF).

21.2.1 Autonomous Systems and Embedded Intelligence

The reliability of a mechanical system depends on its design, the quality of its components, and its maintenance. However, maintenance is typically considered during product design in an ad hoc, ineffective manner, leading to unnecessary life-cycle costs [1].

Maintenance falls into three categories, corrective, preventative and condition-based [2]. Equipment or large assets are often continually monitored with regards to performance and when signals indicate the need for some form of maintenance, action is taken. Preventative maintenance is based upon the mean-time-to-failure (MTTF) value and maintenance is scheduled to occur when required (time-based or usage-based). Occasionally failure can occur outside these bounds and corrective action is taken to repair or maintain the asset.

The traditional image of maintenance is one consisting of Dull/Dangerous or Dirty work. There is also a similarly negative connotation that maintenance is a non-productive element to manufacturing that is only now changing. Maintenance is known to have four characteristics that make undertaking it challenging. Maintenance for a start is often non-uniform because one maintenance sub-operation (removing bolt) may differ depending on the age of the asset (i.e rusted) that makes any pre-programmed autonomous behavior difficult and requires human operators to take uncertain actions outside of their possible knowledge. Second it can be non-standardised, with different actions required depending on the type of asset present, regardless of maintenance sub-operation. Therefore, removing and fitting one type of disk brake pad may vary depending on the maker of the asset (i.e., train undercarriage). Thirdly, maintenance is often irregular, with varying levels of planned and unplanned maintenance which makes determining the right operators or autonomous systems are in place at a given time difficult, particularly if such autonomous maintenance systems can only perform a specific function (inspection, manipulation) and are unable to multitask, something a human is often able to do to meet an increasing workload though only if trained correctly. Finally, maintenance is non-deterministic, because one maintenance sub-operation can change the state of the target system and invoke a different sub-operation from the one that was expected [3].

In the current climate there is a move towards introducing automation and intelligence into a number of products or systems. This is understandable given the benefits self-automation brings, be it the removal of humans from a particular scenario leading to reduced costs or simply their removal from dangerous situations or environments. For example in recent decades there has been an increasing use of robots within manufacturing processes. The speed, power, availability, productivity, and improved accuracy of robots have had significant impact in reducing manufacturing costs while improving production quality. A design for maintenance philosophy that looks to improve the overall maintenance process either through design of the asset itself or to aid in the design of external factors that may contribute to the process in some way, i.e autonomous system solutions is needed [4]. These can incorporate not only characteristics that aid maintenance, particularly with regards to automation, but also exhibit robustness and self-healing attributes.

A large number of maintenance tasks for example simply involve some level of inspection and analysis and if required some form of action to be taken. When considering autonomous maintenance the first step is to think on how this degree of inspection and analysis can either be removed or done in situ within the asset itself. The integration of sensors, for example and other forms of embedded intelligence can provide information to human operators without any form of interaction. Equally they can be used as a tool to aid automation of maintenance by guiding autonomous solutions or provide instruction on any steps that may be needed to complete a particular maintenance task.

21.2.2 Self-healing and Self-repair

To be successful self-healing techniques must be designed to compensate for a wide-variety of failure modes, thus overcoming some of the problems associated with uncertainty. Although specific solutions are not suggested, proposed strategies for developing a self-healing system should not focus on a finite number of underlying causes. Instead the focus should be on how these causes manifest, how they can be detected and ultimately how they can be corrected autonomously. To achieve a self-repairing system, it is clear that the system must have an element of self-awareness. Amor-Segan et al. [5] state that the ultimate aim is to develop a system with "the ability to autonomously predict or detect and diagnose failure conditions, confirm any given diagnosis, and perform appropriate corrective intervention(s)".

A general approach to self-healing applicable to most systems consists of a number of specific steps, 'Cause of Fault', 'Detection of Fault', 'Diagnosis' and 'Corrective Action' [6]. Typically the 0th step, 'Cause of Fault', begins the process, however in the scheme of things the cause is irrelevant in so much as the system should be designed that any behavioral or functional changes are compensated for through further actions. What follows next is the importance of the system to identify the fault, often spatially, with regards to a specific component or region of interest within the system. The 'Detection of Fault' step requires the ability for the system to infer through changes in behavior, function or interpretation of internal or external telemetric data the source of error. The degree of granularity will depend on the amount, accuracy and detail of information retrieved and the systems ability to uses this to come to a confident conclusion of the source fault. Next comes the ability to further analyse the system, and gathered information to bring about a 'Diagnosis'. This is important when looking to initiate some form of corrective action and therefore first be confirmed, to avoid undesirable events such as 'good' components being unnecessarily removed or routed around. Finally 'Corrective Action' will have to be taken to bring back the system behavior back to original specifications, or at the very least mitigate any effects damage may have produced [7–9].

The primary approach to self-repair in electronics involves furnishing the associated circuitry with redundant resources that either mask fault conditions or else remove fault logic, leaving only healthy logic. Redundancy may take three primary forms of spatial, temporal and information redundancy. Spatial redundancy involves the addition of spare physical resources that consume physical space. As depicted in Fig. 21.1, this typically involves copying common resources and connecting them via voting logic. The voter arbitrates between the common resources

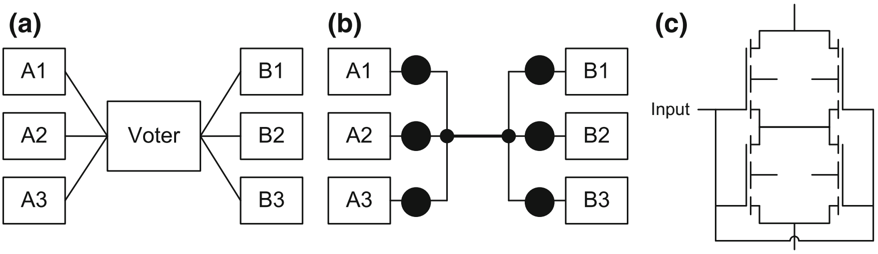


Fig. 21.1 Concept of majority signals in spatial redundancy. **a** Discrete voter. **b** Abstract signal arbitration. **c** Example of signal arbitration within electronic transistor circuit

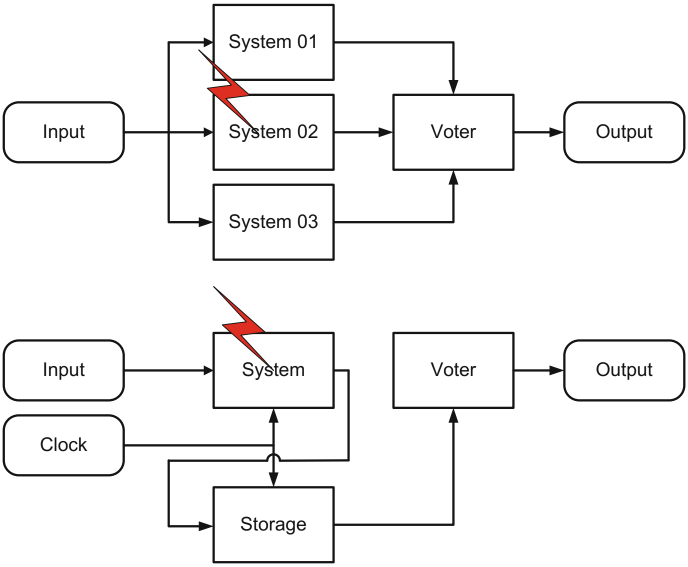


Fig. 21.2 Fault masking within **a** spatial redundancy and **b** temporal redundancy blocks. A temporary fault occurring within either design (represented by the lightning bolt) will not cause output error. However, a persistent fault will compromise the temporal redundancy approach

in order to form a majority output or condition that is passed onward to other circuitry. Thus, fault events may be masked so that their presence does not cause errors within the macro behavior.

Temporal redundancy involves the repeated utilization of resources in time in order to form a majority signal. This is depicted in Fig. 21.2, where the spatially redundant circuit is reformed as a temporally redundant circuit, the primary trade-offs being a reduction in performance and susceptibility to persistent fault conditions.

An example of spatial redundancy applied selectively to a digital NAND gate (a fundamental building block in digital electronics) is depicted in Fig. 21.3. In this

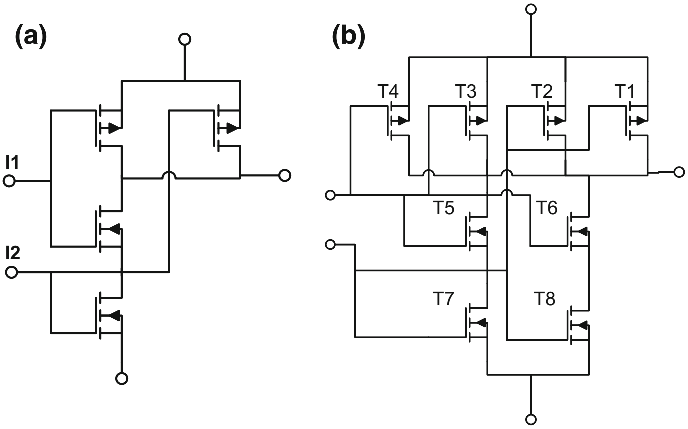


Fig. 21.3 Spatial redundancy strategy targeting ASIC circuitry. **a** Standard NAND gate schematic. **b** Alternative NAND gate schematic supporting selective fault masking and active fault repair trigger

case, the redundancy scheme is chosen to achieve fault masking of specific fault conditions (stuck off faults) while creating a current trigger in the event of stuck on faults [10, 11]. The move towards utilizing redundant resources for self-repair is a complex one due to the need for active detect and repair. However, there are further examples that operate at the transistor level [12].

Besides current off-the-shelf ASICs, future nanoscale devices are fast approaching and will require fresh approaches to fault tolerance and repair. They will be able to support new levels of spatial density but there is debate over the best strategies for achieving resilience [13]. In comparison to contemporary digital electronics design, cell-based designs include bio-inspired evolvable hardware [14] that mimics many properties of bio-cellular organisms including simple inter-cellular actions (nearest neighbor communications), cell states governed by DNA instructions (look up tables) and cellular homogeneity (massive arrays of identical cells). Though this, cellular electronics gains the desirable characteristics of self-organization and adaptation including the property of dependability [15]. The resulting hardware, governing rule sets and software configuration layers are however complex in comparison to standard ASIC design, although there are examples where resource and rule sets have been minimized where possible. Two examples of this are the Plastic Cell Architecture [16] and convergent cellular automata [17], both of which attempt to simplify cell complexity at the expense of repair capability.

21.2.3 Through-Life Engineering Services and No-Fault Found

All systems are susceptible to faults and failures during service, but those faults which occur and result in the No Fault Found (NFF) phenomenon are often unanticipated during the design, sometimes because they occur within acceptable operating tolerances. NFF can be described as a fault whose root cause cannot be found. Therefore there are subsequent inherent diagnostic difficulties. A number of mitigation strategies to combat NFF have been elaborated in literature [18–20], but most are purely that, mitigation for when a potential NFF related failure occurs—such as incorporating specialized test equipment. In truth though, the most significant solution of NFF eradication would be to design and manufacture systems that are increasingly immune to those unanticipated failures that result in NFF. This requires enhanced systems understanding, improvements to the actual design integrity and most importantly a robust mechanism for translating in-service failure (and NFF) knowledge directly back into the design process.

21.2.3.1 In-Service Feedback for TES

NFF events encompass a whole range of products in service, many of which are made up of legacy systems with well-defined operational support practices. Disregarding the fact that the root cause of a NFF will begin with the failure of a component (or unpredictable intermittent faults, which may be part of an inherent design flaw), the end result is a diagnostic failure—that is the maintenance services procedures, equipment, testing capability and guidelines for that equipment was inadequate to isolate the problem. To reduce the NFF event rate for in-service equipment, the conditions under which NFF problems occur need to be considered in depth and investigations should focus on the following areas:

- Failure Knowledge Bases, novel FMECA tools and troubleshooting guides specific for NFF to improve diagnostic success rates.
- Research to pinpoint where in the maintenance process is NFF occurring, for example at a particular maintenance line, testing station, or under specific testing equipment.
- Development of assessment tools to assess maintenance capability/effectiveness which may include:
- Introduction of integrity testing as complimentary to standard ATE (functional) testing procedures.

Although when we talk about influencing (or modifying) a complex system's design for the specific purpose of reducing the impact of NFF, designers need to be clear on what actually needs to be changed in the design. For example, does the system need to be completely redesigned with the aim of increased robustness to make it more fault tolerant? This may have a negative impact on material choices or weight/size restrictions. Or should they change the design so that any faults manifested on the system are accurately detected and located with confidence?

It is the latter of these two questions that would focus on design for testability (DFT). DFT is not a new area of research and it is well accepted that DFT could substantially reduce the effort and cost of testing and supporting a product. However, what is not being done is quantifying the burden of NFF, identifying the root causes, understanding the impact of NFF related faults on coupled systems and feeding this back into the design process with mitigation of NFF becoming a serious DFT goal. The key challenges to address here are:

- Development of design guidelines and standards to improve system designs which incorporate the reduction of NFF as a design goal for improved testability.
- Research into the relationships between system design characteristics and NFF related attributes such as rate of false alarms, fraction of faults isolated to improve design for testability.
- Modelling of complex interactions between system/subsystem/components and their physics of failure.
- Modelling of intermittent failures (identified as the number 1 root cause of NFF) from a fundamental perspective including standardised testing equipment and procedures.
- Developments of a NFF burden/rate predictor for new designs or NFF trending process for in-service systems.
- NFF specific maintenance cost models for design justification.
- In-service monitoring and feedback into design and manufacture.

The fact is that technicians around the world are discovering novel causes of failures on a daily basis. These are being labelled as novel because they have never before been experienced or observed for that particular system, and certainly were not anticipated during the design phase. The reason many failures are not predicted during the design stage is because the system is expected to be operating within a specific set of operational and environmental envelopes a breach of which would signal a predictable failure. However, these novel failures, occur in-service and within the designed operational tolerances making them unpredictable and difficult to diagnose and usually resulting in NFF events.

21.3 Case Studies

21.3.1 Self-healing Electronics

Most electronics self-repair strategies have focused on application specific integrated circuit (ASIC) design that includes the popular field programmable gate arrays (FPGAs). The general goal is to combine fault masking and active mitigation strategies on flexible programmable platforms that support online and offline hardware reconfiguration [21]. However, conventional software-configurable processors must also operate reliably in the presence of faults and their fixed-architecture is therefore enhanced using fault masking strategies involving spatial and information redundancy [22].

Perhaps the most tried and tested reconfiguration strategies involve radiation hardening for space applications. This is due to increased understanding of single event upsets (SEUs) that occur within the Van Allen belt as well as the long-term survivability of deep space and inter-planetary missions. Traditional radiation hardening methods depend upon radiation absorbing materials that add weight and cost. Instead, self-repair through functional redundancy has become desirable, especially when strategies are supported by off the shelf electronics. Prediction and quantification of survivability is a current challenge involving complex reliability modelling to understand the efficacy of redundant resources [23] as well practical experimental validation platforms [24].

A particular example that utilizes the reconfigurability of FPGAs is the Self-testing ARea (STAR) strategy [25] that relies upon creating localized, dynamically changing areas within the active FPGA fabric that quarantine the logic within and perform detailed self-test and repair algorithms. This proceeds while the rest of the FPGA fabric is functioning as normal. The principle is depicted in Fig. 21.4 although this strategy approaches the ideal of complete online self-test

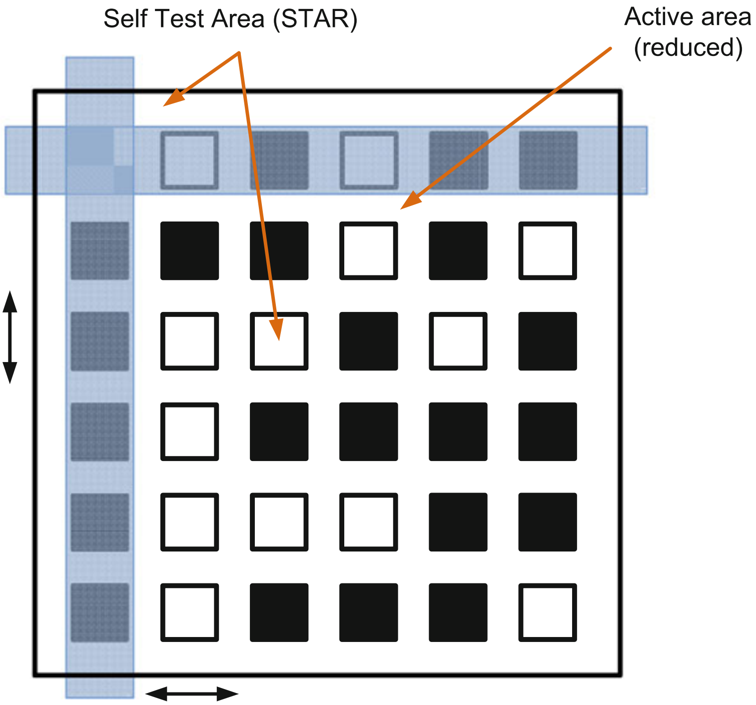


Fig. 21.4 Depiction of the STAR self-repair method for FPGAs

and repair within electronics, there are a number of tradeoffs that typify the challenges of zero-maintenance in general:

- The STAR overhead is considerable and includes the roving configuration logic, test and reconfiguration logic and coordination logic that typically takes the form of an external processor.
- Fault response time is dependent upon the time taken to perform a full test and repair sweep. In [26], Parris estimates the typical sweep time to be the order of 1 s for a modest FPGA device. However, the estimation also predicts sweep times in excess of 17 s for larder state of the art FPGAs. This may represent an unacceptable latency period for some applications.
- Overall throughput of the device is reduced due to the spatial redundancy overhead. In the example given in [25], this amounted to 16% reduction that would impact overall application performance. Given that zero-maintenance extends to applications where performance versus cost is crucial (e.g., the automotive industry), performance throughput represents an additional design overhead that must be factored.

A further complexity is the need for verification of fault repair circuitry of any self-repair strategy. This presents unique challenges, since the current state of the circuit depends on the history of fault events and thus becomes difficult to predict. To help ascertain valid behavior, deterministic fault injection and testing may be performed [27] to quantify the circuit response to specific fault conditions and complemented by probabilistic models where possible [23].

21.3.2 No-Fault-Found and the Failure of Design

One key area that is related to system's design and NFF that is often overlooked is how that system is utilized by the user. For example, is the user operating the system incorrectly because of inadequate training or has the system been designed without the end user in mind? This design issue is a major contributor to the NFF issue across multiple industries. To highlight the case we can consider the consumer electronics industry where the annual bill for NFF returns is in the billions of dollars.

The major impact area for operators due to no fault found device returns is that when customers buy a smartphone, it's increasingly likely that it is a replacement for an existing device. Their expectation is that the new device will perform at least as fast as their old device and also other devices owned by friends and family. If those expectations aren't met because that particular combination of device, application and OTT suffers from wasted data, the customer will return the device as "faulty". The operator then spends time and money testing the device without discovering any faults, and has to resell the device as "refurbished" at a lower margin. This increases their support costs and makes the device less profitable for them.

If a device is not functioning as expected, even with a user error fault, the reason for not obtaining the expected functionality must be identified. One of the underlying reasons for devices being returned as NFF is that the user often has the device configured incorrectly, has misunderstood the device's capabilities or functionality, or there is an underlying hardware/software design problem that is having a secondary 'fault' affect—but that design root cause is not obvious. The user will be required to often contact a service representative who will talk them through a troubleshooting process; however, this can create a frustrated customer when the help and advice provided is not solving the issue. The service representative is not aware of how the device is currently being used and in what conditions, or if the customer is providing the correct information—although this could be rectified through real-time analysis of available data. The cost of service representatives could also be significantly reduced if this real-time analysis for use in device troubleshooting could be delivered directly to the device, allowing the user to identify the nature of the problem.

There is no doubt that the very nature of the NFF problem has its roots firmly embedded within inadequacies present in the design process. This is in no way laying the NFF blame on designers, they will design to the required specifications which almost certainly will not include any reference to NFF—in fact most may not even be aware of any in-service issues that could enhance their designs. In order to eradicate NFF at the design stage this needs to change. Designers need to be provided with information and knowledge captured in the field and more emphasis placed on improved predictability of system usage and operating environments in order to reduce the probability of unanticipated faults occurring. This is no easy task and many of the challenges highlighted in this chapter still remain far from resolved. In addition to this, designers need to be turning their attentions to enhancing the testability of systems. This will ensure that access to test points is easy, appropriate test equipment for the task is identified and the overall test coverage of the system is enhanced. If you cannot test a system then you cannot diagnose it, and NFF will always prevail. Finally, the often overlooked contributor to NFF is how the interaction between the user and system is managed. Systems should always be designed with the user in mind with full training and support to avoid confusion and incorrect operation resulting in perceived faults that again lead to a category of NFF.

21.4 Design for Zero-Maintenance

Systems fail inevitably, in particular complex systems that work across long time scales and within harsh environments. There is no getting around this fact, and as a result mitigation strategies have to be put into place in order return the system to preferably an optimal state. These strategies cover a wide range of fields, services and technical solutions. The most typical comes in the form of maintenance, with processes and actions designed to return function, reduce further degradation and preemptively look to halt future failure. Design also plays a significant role, with a complex struggle in designing solutions that meet cost and performance requirements but also are resilient to damage and robust enough to perform to specifications over the lifetime of the system. The design choices made ultimately affect the performance of the system over its lifetime and how much maintenance is required to sustain this level of performance. The probability of failure of the system or components within the system is naturally also strongly correlated with their design.

Therefore, it is crucial that at the design stage all possible sources or events that lead to failure are understood for the targeted solution. It is also important to understand what the maintenance requirements will be for the targeted design so informed decisions can be made so as to improve its maintainability and resilience to failure. Figure 21.5 outlines a process wheel for the designing in of zero maintenance strategies. The challenge for the designer begins with an understanding of the initial problem or set of problems they wish to overcome.

The first step is to identify the requirements of the system to maintain its function over its life cycle. This is in relation to known faults and possible failure modes, and any processes for maintenance that are necessary for the system to maintain its function to original specifications. It is likely that future design choices further own in the design cycle will introduce additional requirements, however for now it can be ignored. Once the designer has available to him a list of requirements they can begin to reassess which of these requirements can be removed or avoided through a simple design change. In step 2 a designer can begin to prune away certain requirements. It may not be necessary to perform certain inspection practices

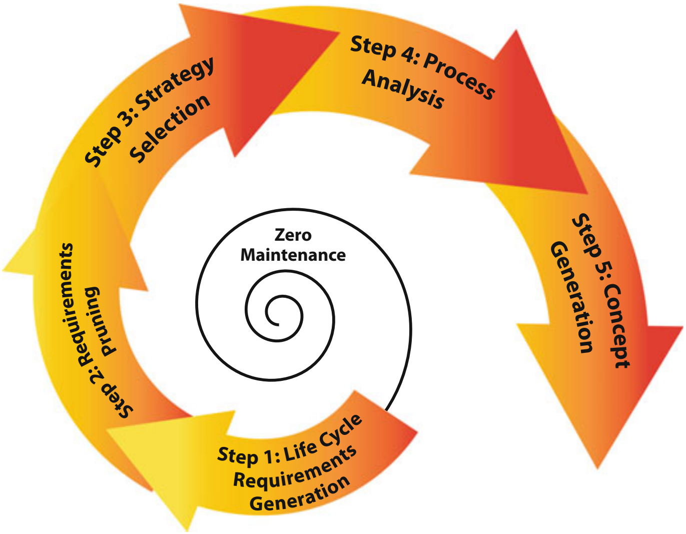


Fig. 21.5 Design for zero-maintenance process wheel

for example if a sensor can replace the role of a human engineer. Careful consideration should occur during this step as it can greatly reduce the number of mitigation strategies integrated into the system. Step 3 forces the designer to make a decision on what direction they wish to take with regards to particular mitigation strategy, whether it is some form of self-healing or repair, or whether outside agents in the form of autonomous systems will provide a solution. Often this will be dictated by the particular failure mode or source of fault, and the component within the system it is related to. Once a strategy has been decided the next step is an analysis of the likely fault process or in the case of maintenance the steps required performing the specific task. Understanding the functional requirements necessary to perform a specific strategy can allow designers to identify alternative design choices to aid the fault recovery or maintenance process. The use of autonomous solutions for example mobile robots can be greatly aided if thought is given into how they may undertake certain maintenance or fault reduction tasks, for example inspection can be helped if diagnostic information is easy to access, or maintenance sites that do not require manipulation to open. Step 5 brings us to concept generation, where for each particular lifecycle requirement, the chosen strategy and functional requirements a designer can look to develop a number of solutions. Finally the designer has to look at the integration of all solutions into a single system and evaluate how such strategies may or may not affect each other. This may result in a different strategy or concept being pursued, or alternatively it may help identify shared functional requirements between mitigation strategies.

When looking to design in electronics self-repair capability its impact is felt not only in the electronic modules, but also the wider system (or sub-systems) given the pervasiveness of electronic and electrical circuits within modern engineering products. In the most extreme cases this brings survivability concepts into the fray across many design levels as shown in Fig. 21.6. Key metrics become fault capacity; latency and repair time become central in the role of electronic sub-systems to assist in securing continued operation in the presence of failures.

With respect to the above example, the performance metrics that need to be considered in design for self-repair are summarized in Table 21.1. Importantly, all such metrics must be considered at the various design stages and fully understood with respect to the strategies under consideration.

Not included in this particular design wheel is an important topic discussed previously, identifying design choices linked with through-life engineering services and their impact. During manufacturing the impact that process variations have on final product quality may cause unanticipated system failures and, in order to address the root causes of such service failures, it is necessary to develop analytical methods based on inter-loop modelling which integrates data from different phases of lifecycle with product and process models. Such a method exists [28] and is summarised below.

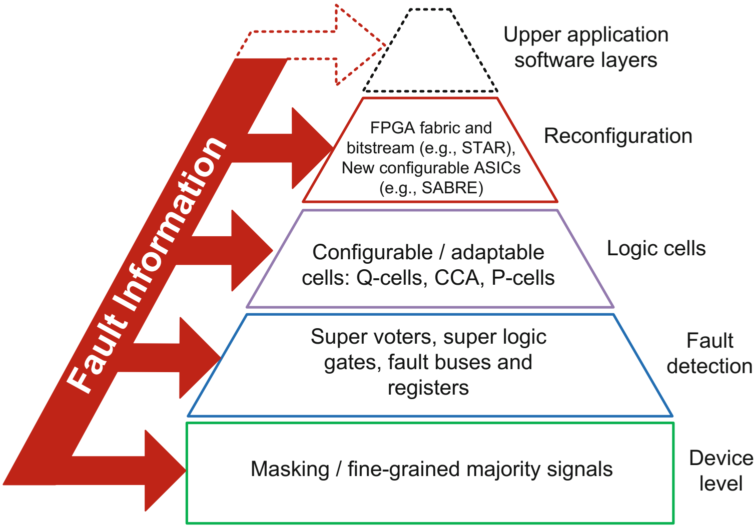


Fig. 21.6 Example hierarchical approach for self-repair focusing on electronic design

Table 21.1 Important performance metrics pertaining to self-repair capability in electronic systems

	Metric	Description
Response	Fault coverage	Refers to fractional area of overall circuit that protected and diversity of faults that can be handled
	Fault granularity	Minimum design layer at which faults can be detected/addressed
	Fault capacity	Number of remaining faults that can be sustained by mitigation strategy
	Performance reduction	Loss of application performance due to self-maintenance operations
	Latency	Timed required for recovery
Resources	Resource overhead	Number of additional components needed over and above basic design
	Resource re-use	Achieving efficient consumption of redundant resources (active methods)
	Energy usage	Energy consumed during recovery and consumed by additional resources overall
Diagnostic	Reporting	Discrimination and reporting of fault events at multiple system levels
	Remaining lifetime	Indication of remaining operational hours after which faults will no longer be handled
	Logging	Capacity to store a log of fault history and classification

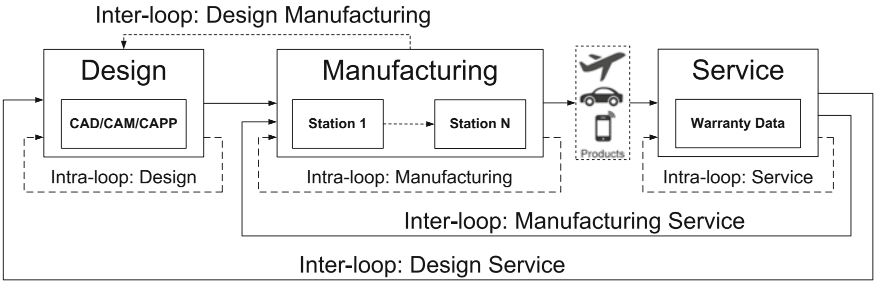


Fig. 21.7 Framework for a closed loop lifecycle model [28]

The loops of a self-resilient production system are classified as intra-loop and inter-loop based on the availability of data from same or different phases of the product lifecycle respectively. Figure 21.7 shows the closed-loop framework. The intra loop refers to integration of data with product and process models from the same phase of the product lifecycle such as SPC that uses manufacturing data for monitoring purposes. The inter-loop refers to integration of data with product or process models obtained from more than one phase of the lifecycle such as addressing service failures.

Within the design phase, product simulation generates data on design parameters that satisfy a set of pre-defined functional requirements. Design changes and optimization is then achieved by modelling the relationship between critical design parameters, functional requirements and critical process variables. During the manufacturing phase, the intra-loop consists of continuous data on the design parameters and process variables obtained using in-line and/or off-line measurements of products and processes during production. The intra-loop in manufacturing is used to address out-of tolerance failures. The monitoring capability can be further integrated with process models to enhance the intra-loop capability of the production systems for fault diagnosis and adjustments. An intra-loop in service consists of warranty data and failure data which are analysed to send feedback to OEMs for setting economic warranty reimbursements to customers, estimating field reliability of products and changing design to address service failures. Warranty data is also used to improve performance of service centres by generating pre-alerting rules to diagnose product failures from customer complaints.

The Design-Manufacturing inter-loop integrates information from manufacturing with design to evaluate and improve diagnosability. By integrating manufacturing and service information together, design parameters can be defined to identify and isolate in-tolerance fault regions. In the Manufacturing-Service inter-loop, the Functional manufacturing and service information is integrated to identify and isolate in-tolerance fault regions. For the Design-Service inter-loop, there is the need for an analytical method capable of identifying root causes of service failures by integrating warranty failures with design models.

References

1. Hernandez G, Seepersad CC, Mistree F (2002) Designing for maintenance: a game theoretic approach. Eng Optim 34:561–577
2. Mobley RK (2002) An introduction to predictive maintenance. Butter-Heinemann, pp 2–5
3. Akrout H, Anson D, Bianchini G, Neveur A, Trinel C, Farnsworth M, Tomiyama T (2013) Maintenance task classification: towards automated robotic maintenance for industry. Procedia CIRP 11:367–372
4. Farnsworth M, Tomiyama T (2014) Capturing, classification and concept generation for automated maintenance tasks. CIRP Ann Manuf Technol. http://dx.doi.org/10.1016/j.cirp.2014.03.093
5. Amor-Segan ML, McMurran R, Dhadyalla G, Jones RP (2007) Towards the self-healing vehicle. Automotive electronics. In: 3rd institution of engineering and technology conference. IET, pp 1–7
6. Farnsworth M, Bell C, Tomiyama T, Khan S (2014) Autonomous maintenance for through-life engineering. In: Redding L, Roy R (eds) Through-life engineering services, decision engineering. doi:10.1007/978-3-319-12111-6_23
7. Farnsworth M, Tiwari A, Dorey R (2014) Modelling, simulation and optimisation of a piezoelectric energy harvester. Procedia CIRP 22:142–147
8. Farnsworth M, Tiwari A (2015) Modelling, simulation and analysis of a self-healing energy harvester. Procedia CIRP 38:271–276
9. Bell C, Farnsworth M, Knowles J, Tiwari A (2015) Self-repairing design process applied to a 4-bar linkage mechanism. Proc Inst Mech Eng Part B: J Eng Manuf
10. Schiefer P, *McWilliam R, *Purvis A (2014) Fault tolerant quadded logic cell structure with built-in adaptive time redundancy. Procedia CIRP 22:127–31
11. McWilliam R, Schiefer P, Purvis A (2015) Experimental validation of a resilient electronic logic design with autonomous fault discrimination/masking. Procedia CIRP
12. Koal T, Ulbricht M, Vierhaus HT (2013) Virtual TMR schemes combining fault tolerance and self repair. In: 2013 Euromicro conference on digital system design (DSD), pp 235–242
13. Han J, Leung E, Liu L, Lombardi F (2014) A fault-tolerant technique using quadded logic and quadded transistors. IEEE Trans Very Large Scale Integr VLSI Syst PP(99):1–1
14. Tyrrell AM (2016) Fault tolerant applications. In: Evolvable hardware [Internet]. Springer, Berlin [cited 2016 Apr 22], pp 191–207. (Natural Computing Series). http://link.springer.com/chapter/10.1007/978-3-662-44616-4_7
15. Tyrrell AM, Greensted AJ (2013) Evolving dependability. J Emerg Technol Comput Syst [Internet] [cited 2013 Feb 20] 3(2). http://doi.acm.org/10.1145/1265949.1265953
16. Nagami K, Oguri K, Shiozawa T, Ito H, Konishi R (1998) Plastic cell architecture: towards reconfigurable computing for general-purpose. In: IEEE symposium on FPGAs for custom computing machines, 1998 proceedings, pp 68–77
17. McWilliam R, Schiefer P, Purvis A (2015) Creating self-configuring logic with built-in resilience to multiple-upset events. Proc Inst Mech Eng Part B J Eng Manuf [Internet] [cited 2015 Oct 26]. http://pib.sagepub.com/content/early/2015/10/01/0954405415611607
18. Khan S, Phillips P, Jennions I, Hockley C (2014) No fault found events in maintenance engineering part 1: current trends, implications and organizational practices. Reliab Eng Syst Saf 123:183–195. http://dx.doi.org/10.1016/j.ress.2013.11.003
19. Khan S, Phillips P, Hockley C, Jennions I (2014) No fault found events in maintenance engineering part 2: root causes, technical developments and future research. Reliab Eng Syst Saf 123:196–208. http://dx.doi.org/10.1016/j.ress.2013.10.013
20. Soderholm P (2007) A system view of the no fault found (NFF) phenomenon. Reliab Eng Syst Saf 92(1):1–14
21. Cheatham JA, Emmert JM, Baumgart S (2006) A survey of fault tolerant methodologies for FPGAs. ACM Trans Autom Electron Syst. 11(2):501–533

22. Patel A, Prakash K (2010) Fault-tolerant features of modern processors—a case study. University of Wisconsin-Madison
23. Liu Y, Trivedi KS (2006) Survivability quantification: the analytical modeling approach. Int J Perform Eng 2(1):29–44
24. Kastensmidt FL, Reis R (2010) Fault-tolerance techniques for SRAM-Based FPGAs. Softcover reprint of hardcover 1st edn. Springer, 198 p
25. Emmert J, Stroud C, Skaggs B, Abramovici M Dynamic fault tolerance in FPGAs via partial reconfiguration. IEEE Comput Soc [cited 2012 Apr 25], pp 165–174. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=903403&tag=1
26. Parris MG, Sharma CA, Demara RF (2011) Progress in autonomous fault recovery of field programmable gate arrays. ACM Comput Surv 43(4):31:1–31:30
27. Validation resilient logic units by automated fault injection–solutions—national instruments [Internet] [cited 2016 Jun 26]. http://sine.ni.com/cs/app/doc/p/id/cs-17146
28. Pal A, Franciosa P, Ceglarek D (2014) Root cause analysis of product service failures in design—a closed-loop lifecycle model approach. Procedia CIRP 21:165–170
29. de Novaes Kucinskis F, Ferreira MGV (2010) Taking the ECSS autonomy concepts one step further. In: SpaceOps 2010 conference "Delivering on the Dream" Hosted by NASA Mars [Internet] [cited 2015 Apr 14], pp 25–30. http://arc.aiaa.org/doi/pdf/10.2514/6.2010-2364
30. Lee J, Ni J, Djurdjanovic D, Qiu H, Liao H (2006) Intelligent prognostics tools and e-maintenance. Comput Ind 57(6):476–489

Chapter 22
Graph-Based Model for Context-Aware Maintenance Assistance with Augmented Reality and 3D Visualization

Michael Abramovici, Mario Wolf and Matthias Neges

Abstract The benefits of augmented reality applications in maintenance are widely known and since high-performance smart devices are the common standard for mobile devices, the actual preconditions for the usage of such applications seem promising. Problems emerge whenever service is to be conducted in an area of banned photography. Using a smart device with a camera is either simply not allowed, or the internal camera has to be pasted over to allow entrance into the restricted areas. Either way, the technician still relies on his maintenance assistant system to provide useful information if he does not want to go back to pen and paper. In this article a concept is elaborated that offers context-sensitive guidance, a highly dynamic data model and different views, depending on the availability of an internal camera and/or restrictions of the work environment. The approach presented was implemented and validated under laboratory conditions with a complex hydraulic system as a demonstrator machine. The prototype will be the foundation of an industrial case study concerning the combination of IoT enabled machinery and smart devices in maintenance later this year.

22.1 Introduction

Augmented and virtual reality are current trends in the development of maintenance assistant systems (cf. [1, 2]), while 3D computer graphics are state of the art in product development and visualization. As the benefits in presentation and interaction are widely known these technologies are used for innovative solutions for mobile applications [3]. Schlick et al. [4] analyzed training success based on the instructional media and found that graphical instructions are superior to textual ones whenever the repetition rate is low. As mobile devices (smart phones, tablets) have

---

M. Abramovici · M. Wolf (✉) · M. Neges
Ruhr-University Bochum, Universitätsstraße 150, 44780 Bochum, Germany
e-mail: mario.wolf@itm.rub.de

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_22

fully saturated the market [5] throughout the last decade, augmented reality applications are ready to be used on a large scale.

Augmented reality is defined [6] as overlaying reality with additional virtual content and as such relies on camera footage, which disqualifies the whole area of augmented reality applications for usage in company sites with a ban on photography. Especially ancillary industries are often bound by contract to facilitate such zones to prevent industrial espionage. The question arises whether the camera-reliant augmented reality application's content could be used for a camera-independent mode without reducing the application's benefits.

Due to networking capability of products in the Internet of Things (IoT) and the accompanying availability of data, products can not only operate, react and start actions autonomous but even use additional services and evaluate the data received from external sources [7]. This among others offers enormous potential in maintenance of technical machinery, like the improvement of service documentation by combining it with data from the actual machines [8, 9].

A context sensitive support of maintenance with Augmented Reality (AR) for instance reduces worktime by 55% in a complex pipe-systems while at the same time halves the error rate [10]. Augmented Reality applications allow an interactive problem-solving, which enable intuitive work with an interactive guide that is dependent on the actual state of the smart product, rather than static step by step instructions.

Generally, in the context of maintenance assistance, there are certain common features between augmented reality applications and applications that display 3D models to enrich the information for their recipients. Such is the need for the preparation of work instructions, the usage of 3D mockup models instead of full-fledged Computer Aided Design (CAD) models and the need for precise orientation of the afore mentioned models [11].

When taking into account the similar functionality of augmented reality and 3D model displaying applications, a situation-dependent visualization mode should be implementable.

In this paper, we

- determine requirements for a camera-independent mode in an AR application,
- define the planned functionalities,
- describe the theory behind marking technology-based identification, how it can be used for adaptive 3D-Model orientation and
- validate our approach with two use cases in each visualization mode.

The result will be presented as prototypical solution consisting of a back-end service for data acquisition, data analysis and a front-end AR application to support the maintenance technician in the field. Since the potential of the use-phase of Radio Frequency Identification (RFID) and Augmented Reality applications have been treated already [12, 13], this chapter will focus on the dynamic data-model and the multi-functional maintenance assistant.

The first use case includes the visualization of real-time operational data, while the second case includes the actual maintenance assistance with context-sensitive status analysis and indications. To realize context awareness, we use our validated approach on graph-based algorithms [13] to guide the technician after evaluating the current state of the machine, whether it is in regular AR- or camera-independent mode.

22.2 Aims and Requirements

The aim of the presented chapter is to better support technicians on-site through smart usage of existing information and mobile devices, even in zones of banned photography. To achieve that, the range of functions of the components discussed in [13] need to be extended, either to improve shortcomings or to enable new functionalities.

The main goals are to minimize the need for manual inputs, offer real-time information about the machinery at hand in a manner that is quick and easy to grasp. All that should work independently of the availability or legitimacy of a camera on the smart device.

Taking advantage of the Internet of Things (IoT, cf. [14]) the current approach will no longer contain the error-prone (cf. [15]) manual gathering of the operational data but analyzes real-time sensor data directly read from the machine at hand and evaluated trough modules within the IoT platform.

The Goal of expanding the concept is the inclusion of information from various source-systems like Product Lifecycle Management (PLM) and Computerized Management Systems (CMMS). These types of applications usually are designed to be used on a desktop PCs, with some concepts to bring them to mobile devices without enhancing the functionality [16].

Joint with the possibility of being able to analyze real-time sensor data of the momentary machine status, potential arises for dynamic responses of the assistant application in form of indications, hints and warnings displayed in augmented reality [17] or 3D during a maintenance task.

After completing the maintenance task an automated documentation and creation of work performance record is conducted to minimize most of handwritten work by the technical personnel. The following synchronization of mobile application and the graph-based model provides a way of efficient data management and feedback repatriation of data that exceeds the regular handwritten coverage in both coverage and future-proofness. This way machine or component specific data for the manufacturer or contractor can be retraced to enhance attrition calculation or product usage data for the product development of the successor generation (cf. [9, 18]).

The assistance system is dependent on sensor data, so our demonstrator had to be fitted with according devices. Particularly interesting are wireless sensors on basis of Zigbee Standards (IEEE 802.15.4) or Radio Frequency Identification (RFID) ISO 18000-6C, as these are easy to install and configure in a prototypical environment. Theses sensors then must be equipped to send their payload to the cloud-based IoT platform for further processing.

As consequence of the desired functionality as part of the Internet of Things (cf. [19]) an active internet connection is needed for communication of the integration platform in order to access the real-time operational data and make use of the analytics functions.

As further electronic documentations and 3D-perspectives of the concerning components shall be made available, the selection of the possible viewing options is crucial.

To get a grasp on the problem of representing certain information, one must understand the specialties of the different projection modes. As the unrestricted viewing mode shall be augmented reality, the functional range of AR is the reference point. Broll [20] defines augmented reality as the enhancement of reality with artificial digital content, while the interaction is real-time capable and interactive. Furthermore, the enhancement must be continuously adapted to the user's perspective to enable the fusion between digital and real elements. Intuitively, when taking away the camera picture what is left of augmented reality is the digital content alone, which means 3D computer graphics in this context.

On the other hand, following the Milgram's reality-virtuality continuum [21] from augmented reality in the direction of *less reality*, the final destination is virtual reality (Fig. 22.1).

Broll [20] characterizes 3D computer graphics as a purely visual non-immersive presentation, that is not time-sensitive, user independent (exocentric perspective), displaying only a static scene with 2D-Interaction with i.e. Mouse and Keyboard. By contrast, virtual reality applications feature multimodal presentation (visual, acoustic, haptic), real-time interaction, user centered, immersive presentation (egocentric perspective) and 3D-Interaction (hand-, head- and body-movement).

The desired solution for the presented approach shall offer the benefits of augmented reality and virtual reality in regard of real-time capable interaction through movement and gestures, egocentric perspective and information density, while neglecting immersion for ease of use with smart devices without head mounted displays (HMD). As 3D models are available for overlay-display in augmented reality, these models could be used in a camera-less visualization mode if the maintenance is to be done in an area of prohibited camera use.

Fig. 22.1 Adapted reality-virtuality continuum based on Milgram et al. [21]

22.3 Concept

The overall approach is subdivided into the three modules *Cloud Data Sourcing*, *Cloud Data Handling* and *On-site Maintenance Assistant* to represent each functional unit and labeling the distribution pattern. Each module is then further subdivided into two layers to separate distinct functionalities within said functional units. Conceptually the flow of payload data, meaning data that is required to enable the technician on-site to get the actual information he needs, is strictly bottom-up (cf. Fig. 22.2). As the technician is not only the recipient of information, but also the author, the concept must therefore allow feedback data to flow back into the responsible components for later use.

The three main modules will be explained further in the following subsections.

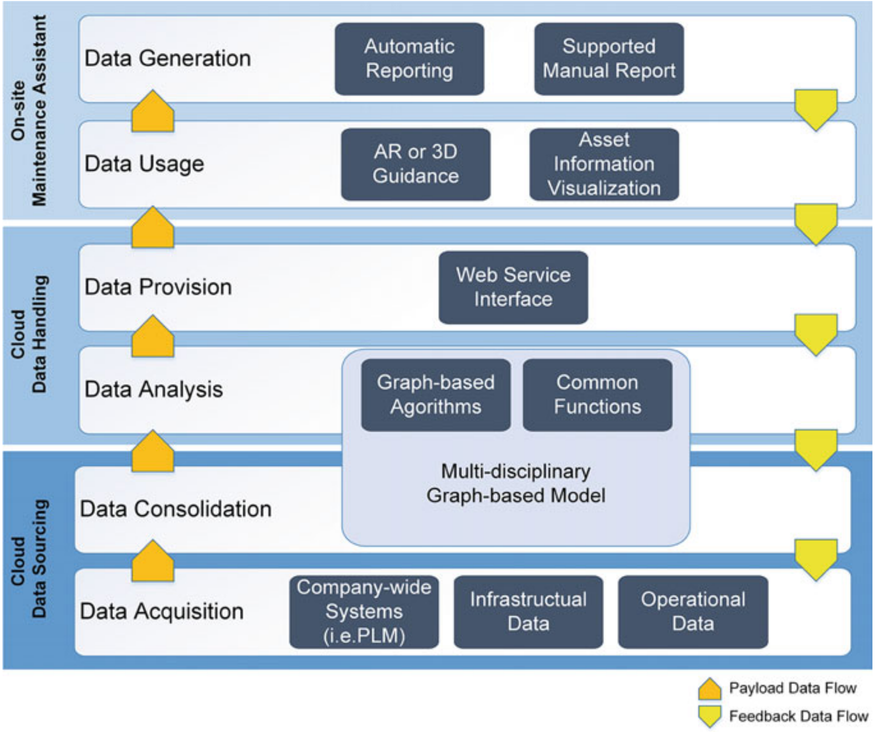


Fig. 22.2 Conceptual system architecture

22.3.1 Data Sourcing

In this module, data sources for processing in the multi-disciplinary graph-based model can be found, as well as the mechanisms for bringing the data into an evaluable form.

The *Data Acquisition* layer gathers data particularly from PLM and CMMS as this information is needed for the technician's work on-site. Along with the data extraction from the company-wide systems the interfaces are used to backchannel feedback information from the maintenance processes into the graph-based model and from there into the company-wide systems thus ensuring that the company-wide systems possess the leading data basis and are up to date.

The operational data is not only gathered from sensors of individual machines but also from various other services like location-based services (i.e. weather forecast), ambient data from the facility, availability data for surrounding systems or similar context-relevant data. As the administered amount is high, consolidation and filtering of the data is needed, before an actual interpretation for the use through a technician.

Haberfellner and Daenzer [22] cite that models as a depiction of reality must be convincing enough in regards to the situation and problem position. This means that in all deliberations the question of functionality and problem relevance is to be asked. To answer the question of functionality and relevance Harrison [23] cites the strength of graph databases lie especially on the use cases in which the correlation of elements is more important than the elements themselves.

While company-wide systems specialize on administration of inventory data, the multi-disciplinary graph-based model is used as external storage for the maintenance data and information extracted and aggregated from existent systems, to be dynamically evaluable for service technicians. To control the complexity of the model creation (cf. [24]) domains which are involved in the setup of the model are limited to product structure, electric, pneumatic and hydraulic networks, as well as task management.

It should be clarified that the functionalities of this approach are based on the considerations of Harrison, Haberfellner and Daenzer, which lead to the following questions:

- Which *things* are relevant to the activity? Which *things* are affected by the planned activity?
- Do *connections* exist between those things?
- Are there *rules* to be identified in order to evaluate connections between things in an algorithmic fashion?

If this basic idea is transferred into the world of graph theory things are consequently displayable as vertices. These vertices represent certain entities with all their attributes. Connections between the individual things will be mapped with edges. Edges in graphs can have attributes on their own. Constraints ensure the

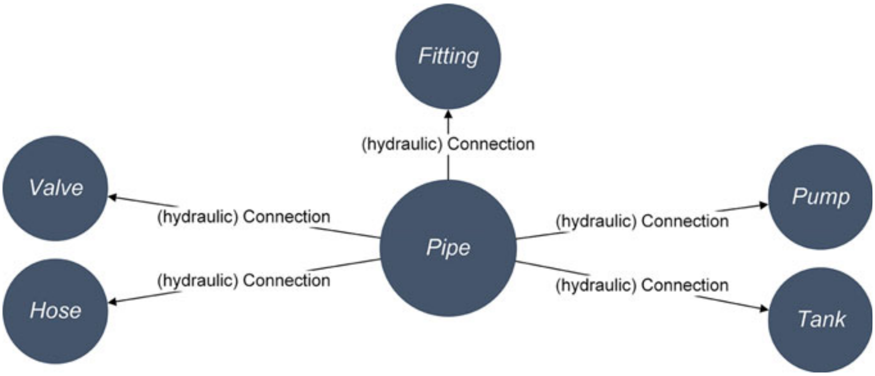


Fig. 22.3 Exemplary valid outgoing connections between hydraulic components

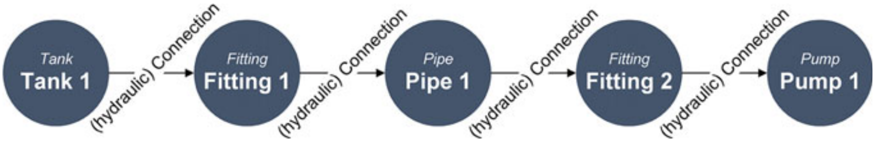


Fig. 22.4 Exemplary network of distinct hydraulic components

necessary rules for validity of connections, so that only certain types of vertices may have certain type of in- and outgoing edges.

An example of implementing this order is displayed in Fig. 22.3. It is to show the valid outgoing connections between various hydraulic components from the vertex of type *Pipe* and elucidate the appropriate rules by reference on a clear example.

Following the pattern vertices of type *Pipe* are solely entitled to outgoing edges of type (*hydraulic) Connection* to the types *Valve, Hose, Fitting, Pump* or *Tank* as their target.

The classification of relevant components of maintenance focuses on product characteristics rather than on product attributes like stated in [24]. Particularly the characteristics which can be identified quickly and safely on-site by the technician should be chosen, such as allocation of a certain functional groups, for instance types like seen in Fig. 22.3 and similar discrete physical object types. Types in this context generally can be understood as an equivalent of classes in object-oriented programming. This means once a class is designed, you can create discrete, distinguishable objects of the same pattern, but with different characteristics, i.e. two pipes with different lengths. An example for valid instantiation of said classes can be seen in Fig. 22.4.

The next step in the process of building a product-characteristic-based graph presentation consists of the modelling of classes and their subsumption in hierarchies. These form the following perspectives of the data:

- Hierarchy of the vertex classes
- Hierarchy of the edge classes
- Constraints between vertices and edges

To ensure a repeatable identification capability the classification of components was arranged according to the component's functional characteristics in context of maintenance. The same procedure was performed on the relations and their corresponding constraints to avoid erroneous entries. Due to these arrangements a combination of an electric component as target of a hydraulic flow is (fortunately) impossible.

These preparations build the basic frame for the graph-based model, which is by design sound in itself and allows considerations of all relevant elements in the maintenance process. Once the data is collected from the source systems, the utilization of said objects of classes of different relevant elements in maintenance can begin and thus *Data Consolidation* is provided.

22.3.2 Data Handling

Once the model is set up utilizing the rules mentioned in the last section it can be used to gain insight about the status of machines at hand. The *Data Analysis* layer denominates the entirety of the provided functions and algorithms to interpret or manipulate the data in the graph-representation. Using a mixture of classic relational query design and graph-based algorithms (cf. [13]) like the Dijkstra or Floyd-Warshall, one is able to find paths between relevant start and end point in the graph network.

For instance, an analyzation can be started by following hydraulic connections between relevant elements of a facility (i.e. a cooling cycle consisting of a multitude of parts), which translates to finding relevant vertices of correct type and following the right classes of edges to form a path.

If basis and goal are known a pathfinding of the graph is possible without additional features of edges and vertices. The addition of such information enables either more precise results or concrete conclusions based on further algorithms.

If the actual representations of functional elements in the graph are conducted as displayed in Fig. 22.5 it becomes obvious how the before mentioned classification can help with solving certain problems by traversing a graph and evaluating the data collected along the found path.

Taking up the afore mentioned example of the wished analyzation of a certain cooling circle, the following question arises:

Question: Are there any current or upcoming problems in cooling cycle X?

In order to answer this question a path through the graph of the cooling cycle X needs to be found via an algorithm. Once it is found the path will be iterated front to back, while on each vertex relevant neighboring vertices connected through selected edges of the right classes are analyzed.

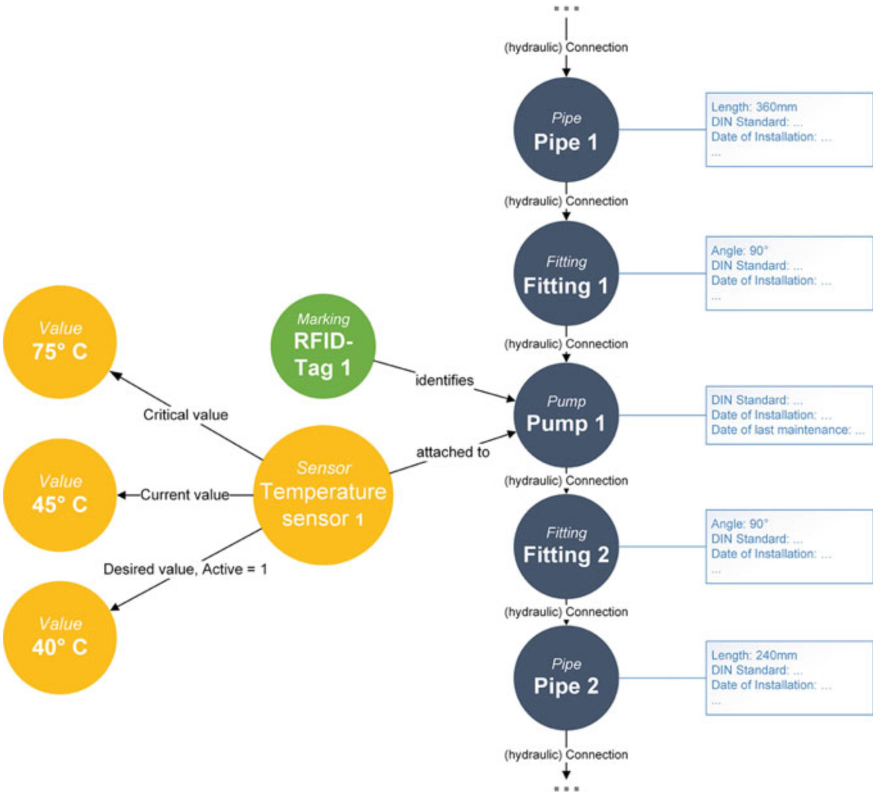


Fig. 22.5 Exemplary excerpt from the graph model

Applied to the example in Fig. 22.5 one can see that each vertex has stored data and connections to other vertices. The algorithm starts off and finds a path that includes the section between *Pipe 1* and *Pipe 2*. After the successful path-finding, the found vertices are iterated and checked against predefined functions to evaluate their current status. *Pump 1* is equipped with the sensor *Temperature Sensor 1* which is connected to value vertices. The values in itself hold data about a certain value and the value's measuring unit. Through this process each sensor is processed individually and offers maximum flexibility. The information about the value's significance is expressed in the type of connection between the sensor and a value. In this example *Temperature Sensor 1* offers information to the system which basically translates to the answer of the initial question:

Answer: The temperature of Pump 1 is higher than desired but within the tolerance range.

To get this process to create work instruction or indications for the technician in the field, it is not necessary to create step-by-step instructions for instructions concerning sensor-equipped elements, but to create *Jobs* with *Tasks* that include

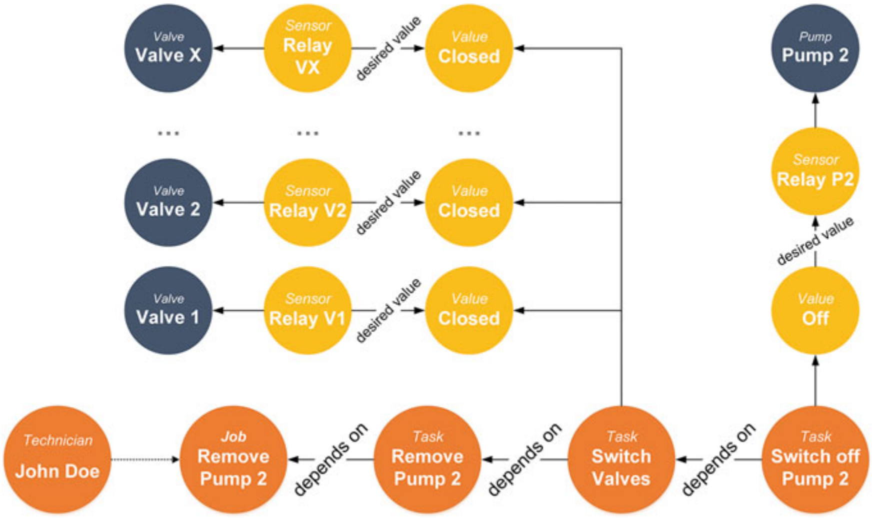


Fig. 22.6 Exemplary job for automated instruction generation

dependencies between the tasks and the desired *Values* of sensors concerned in the planned maintenance process. To elaborate this, Fig. 22.6 shows an example where two tasks are defined with values that need to be in place before the task is completed. The first step in this *Job* to remove *Pump 2* is to switch off the relay controlling the pump. No work instruction would be created if the desired status was equal to the current one. Same goes for the status of valves in the follow-up task. Only for non-closed valves would an instruction be generated.

The desired values which are connected through the according edge class can specify a criticality level based on ISO 26262's hazard analysis and risk assessment. The resulting criticality levels in the concept are defined as follows (S = Severity).

S0: No damage, no functional components affected
S1: Light damage, reduction of efficiency
S2: Hazardous damage, loss of functional components
S3: Catastrophic damage, loss of the safety related components or complete machine

Although it is in the nature of such a concept to capture and evaluate data from multiple (conceptually from an unlimited number of) sensors, we do not implement an actual multi-sensor data fusion. Hall [25] states that a multi-sensor data fusion is characterized by statistical advantage gained by combining same-source data or increased accuracy trough use of multiple types of sensors.

In the current state of this concept, the focus lies on the dynamic evaluation of the captured data, based on functions and algorithms that offer direct evaluation.

As stated in the requirements section, the data providing modules are to be made available as a cloud-solution via internet connection. The *Data Provision* layer thereby consists of a set of web service providing components to communicate with the on-site clients, as well as other systems on a remote location.

22.3.3 Maintenance Assistant

The maintenance assistant application provides the user interface for technicians. The *Maintenance Assistant* is conceived as an application for mobile devices and supports the maintenance personnel on-site with data provided by the *Data Analysis* layer.

From here the technicians can access all offered functions for their specific task. The functionalities of the application can be divided into two categories, first the *Data Usage*, second the *Data Generation*.

The authors already published a description of a graph-based maintenance assistant system for smart devices in [13]. The then developed concept for an augmented reality application must be adjusted to match the new circumstances, for instance the pre-known current machinery status through implementation of IoT components and the confrontation with possible a ban on photography.

In order to use an augmented reality application which obviously is dependent on its camera various arrangements have to be made. Components of an augmented reality application are arranged in the categories *display, tracking and interaction* [26]. The virtual part of the augmented *display* principally consists of digital 3D-models. To increase performance on mobile devices, the general approach is to reduce 3D models to mock-ups, that still feature the same overall geometry, but are far less complex.

Tracking in case of augmented reality denotes location determination of user (or the displaying device) as well as from distinctive characteristics and objects in space. This can be achieved via various procedures which the authors [27] looked upon in an earlier publication. The possibilities of interactions in the AR-environment are versatile and span from marker-based input, motion tracking or simple touch gestures on the display of the smart device [26].

In Chap. 2 the reality-virtuality continuum was discussed along with the possibilities to use AR application components to build a camera-independent 3D application. Starting again at the core components of an augmented reality application (display, tracking and interaction) the guidelines for the development of the 3D mode can get designed.

As already stated, the 3D mock-up models are presented as an overlay over the camera picture within an AR application. The same models can be displayed from within the 3D engine that renders the 3D models with an artificial background instead of a camera picture.

As for the tracking component it is obvious that without the camera and for that regard similar systems like infrared or laser based measurement systems, it is not possible to optically track the relative position between the user/the user's smart device and the concerning machine. For the initial pose a small set of data can be stored right on a RFID tag on the machine to align the model. This data set can be encoded in JSON and stored as regular text on the tag to be used on every RFID-enabled device. An example for such a data set matches the following pattern to provide the application with a unique identification, a human-readable description and the initial roll-pitch-yaw angle:

```
{

“ID”: “1234”,
“Label”: “Pump 1”,
“RPY-Angle”: [15, 0, 90]

}
```

Instead of using the positions of markers or inherent features calculated with the AR-framework, a gyroscope which is usually fitted in every smart device nowadays can be used to keep the 3D model in roughly the right angle to the user after establishing the initial pose.

Interaction will be implemented via gestures in order to steer the 3D-rotation, as is the standard in modern mobile devices. Basically the same principles apply to both viewing modes, so that the *Data Usage* perceived as representation of the guidance is identical from a data perspective.

The concept’s top layer *Data Generation* is responsible for creating the feedback data that is brought back to the source systems. This data may consist of log files of fulfilled tasks, used material, exact timing, involved personnel and parts and manual entries from the technicians.

Implementation targets for the maintenance assistant application are industrial mobile devices, generally in the category of tablets. Given that the tablet at hand is equipped with a camera, the maintenance assistant will be able to switch to either 3D or AR mode at the beginning of a maintenance process.

22.4 Prototypical Implementation and Validation

For the demonstration of validity of the presented approach our validation is centered around the complex hydraulic system (Fig. 22.7) which was also used in the case study in [12].

The hydraulic system has two centrifugal pumps, two tanks and twelve valves. The construction of the hydraulic system allows the fluid to be pumped through both of the two pumps and through each of the two containers, as well as through both pumps in a row. The respective fluid circuit depends on the combination of the respective valve positions.

The hydraulic system was upgraded with newly designed sensors to gather the status of the valves and pumps automatically. Sensor modules were designed and manufactured which gather the lever positions of the valves and so distinguish

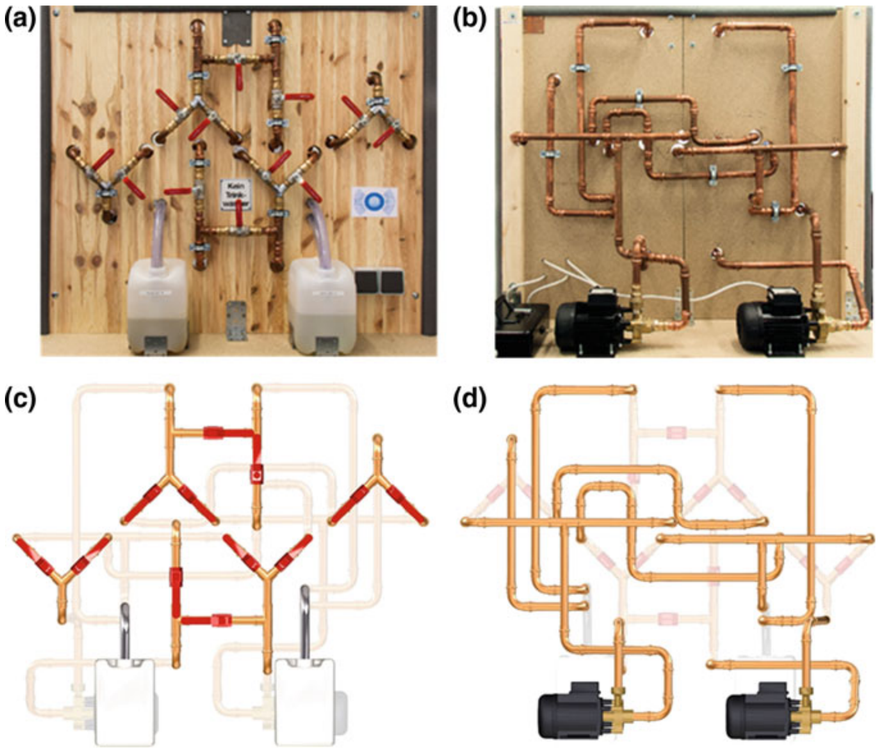


Fig. 22.7 Front (**a**, **c**) and backside (**b**, **d**) of the hydraulic system in both reality and virtuality

between open or closed lever positions as well as capturing the unsafe state in between. Furthermore, the installed pumps have been wired via relays which are switchable and readable by the controlling Raspberry Pi 2 on the basis of self-developed web service-protocol. With this the sensor data of the hydraulic demonstrator is accessible for the data sourcing module via the mentioned interface.

The general structure of the concept (cf. Fig. 22.2) has been preserved in the implementation (cf. Fig. 22.8). For easier understanding the functional units were remodeled after the distribution of functionalities offered by self-developed components as well as commercial software.

The software products used were:

- Data Sourcing
  - Data Acquisition: PTC Windchill, Self-developed components
  - Data Consolidation: PTC ThingWorx, Self-developed components
- Data Handling
  - Data Analysis: PTC ThingWorx, OrientDB, Self-developed components
  - Data Provision: PTC ThingWorx, Self-developed components

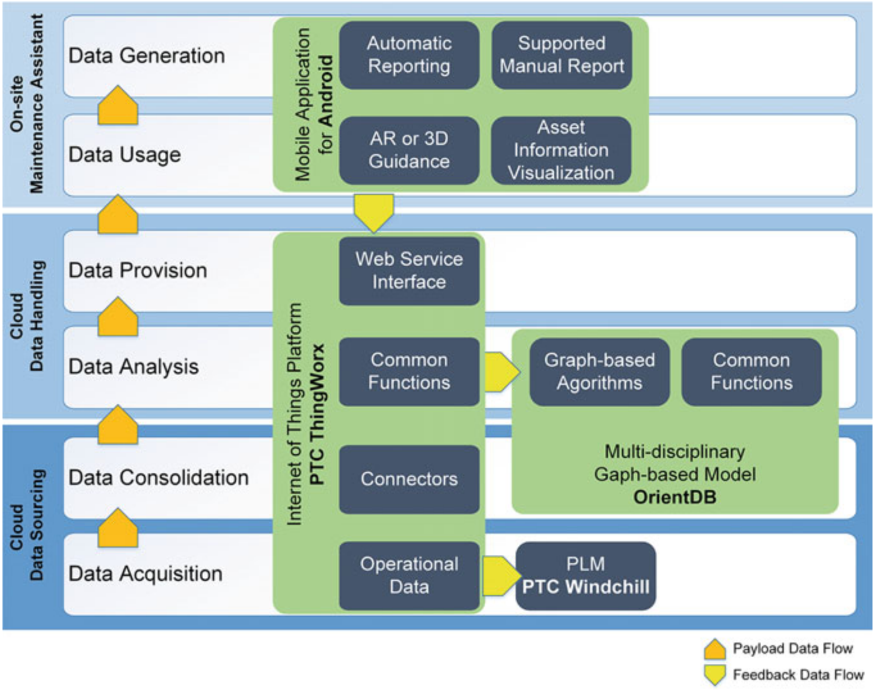


Fig. 22.8 Implemented system architecture

- Maintenance Assistant
  - Data Usage: PTC Vuforia, Google Android, Self-developed components
  - Data Generation: PTC Vuforia, Google Android, Self-developed component

Generally, a loose coupling of components on web services of XML- or JSON basis was implemented in the prototype in order to achieve maximum flexibility. These technologies are generally considered to be feasible for the use in IoT applications [19].

Two major components are relevant in this module of the implementation. First there is PTC Windchill as a source system for 3D models and meta data, secondly there is PTC Thingworx as the central point of communication between the different components and product instance specific data.

The 3D design data was deposited in PTC Windchill for later use. With help of the so called connector, Windchill-data can be accessed on the IoT-platform PTC ThingWorx. In addition, ThingWorx is capable of seizing precast meteorological services and retrieve these for each specific GPS-coordinate provided.

The hydraulic system possesses a digital twin [28] in PTC ThingWorx, which means every set of data referring to the special product instance (operating data and for example spare parts etc.) are administrated here.

OrientDB is a graph database with document oriented store option [28] and in this implementation utilizes predefined a conceptual graph database schema according to the concept described in Sect. 3.2 instead of an automated import from i.e. the product structure. Within the context of the prototypical implementation this work was done by hand in order to examine the functioning and evaluate the dynamic extensibility rather than laying focus on the import process from existing models. The import process and the resulting automated model building will be a subject in future research.

The administration of product instance data takes place in the PTC ThingWorx, while the administration of the semantic and operational data is organized in OrientDB.

The planned functions for the graph traversing and analyzation were deposited in the graph database management system as java script user functions and made accessible for other components via the build in PHP-web service-implementation of OrientDB.

The *Maintenance Assistant* application was implemented with PTC Vuforia and the open source 3D-engine JPCT-ae. As mentioned in the concept, the application supports two modes of usage to present either augmented reality or 3D content.

To show the capability of the approach at hand we prepared two use cases in laboratory environment. The first being the simple monitoring of operational data, the second being the maintenance assistance with context-sensitive indications.

22.4.1 Visualization of Real-Time Operational Data

The first implemented operation mode shows selected operational data or to be more precise analyzation reports based of operational data. The implemented functions for data evaluation have an immediate impact on the options the technician can choose from on-site. The following options were implemented to choose sensors based on

1. their type,
2. their unique identification,
3. the family of parts they are attached to,
4. their general position on the machine or
5. the parts that are relevant in the assigned task.

These modes generally filter the multitude of managed sensors in the graph-based model. As shown in Sect. 3.2 sensors are connected to their specific boundary values individually while they are analyzed according to their edge classification. Further filtering can be applied by selecting a minimum criticality level for the displayed indications. As an example only “*category D: critical condition*” evaluated indications concerning the front side of the hydraulic system

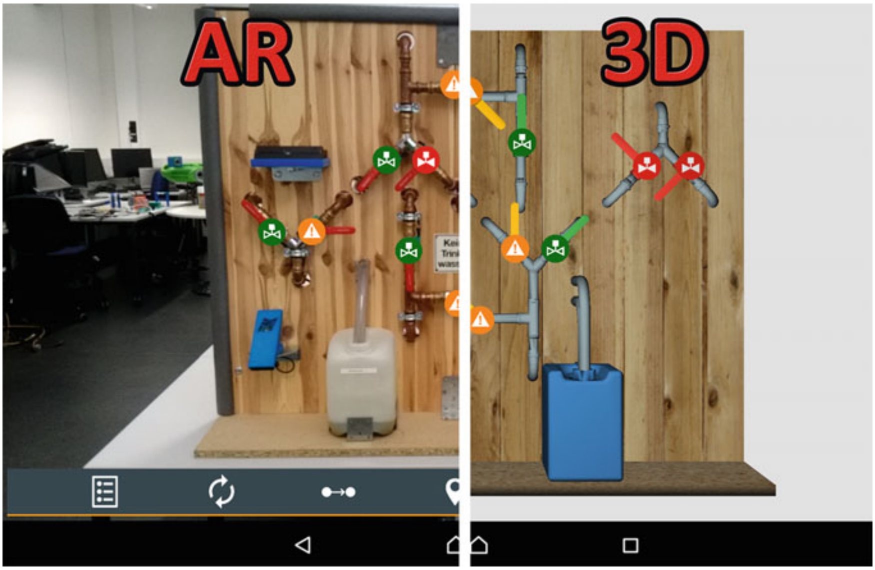


Fig. 22.9 Augmented reality visualization (*left*) and 3D visualization (*right*) for warnings

are shown in Fig. 22.9. The center value on the very top and bottom of the machine are in an unknown state, which causes the functions to report a critical condition.

If a change of state occurs (i.e. by switching a valve, which triggers updates in the graph based model) and exceeds a critical limit in the monitored system, the particular technician immediately gets automated warnings and indications concerning the current context. If the change brings the system into an uncritical state, only minor indicators are updated.

22.4.2 *Context-Sensitive Maintenance Assistance*

The second use case is comparable with the implementation the authors conducted in [13]. The maintenance assistance with context-sensitive work instructions was implemented based on the assumption that *pump 2* needs to be removed (cf. Fig. 22.6) for later replacement. For that matter *Pump 1* needs to handle the flow through both tanks in the system.

This premises induce certain desired conditions which then are evaluated considering the machines current state to generate the required instructions for the technician.

Figure 22.10 show the resulting instructions and indications. The list to the left of the screenshots offers a quick overview on the correct and incorrect valve positions for the working step “switch valves” in Fig. 22.6 in reference to the systems status at that time.

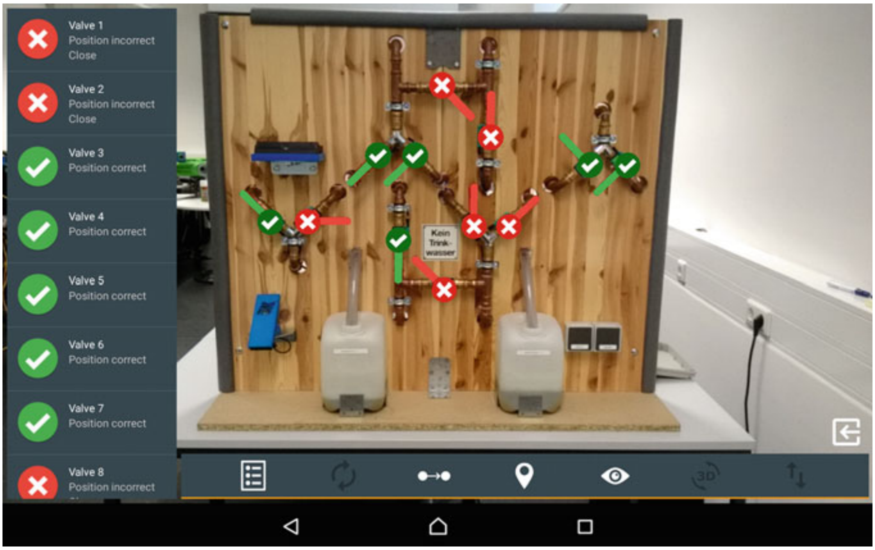


Fig. 22.10 Augmented reality visualization for maintenance guidance

It is easy to see that while the augmented reality mode (Fig. 22.9, Left side) offers the intuitive viewing of instructions and machinery alike, the amount of information shown is identical to that of the camera-less 3D mode (Fig. 22.9, right side).

22.5 Conclusion and Future Work

The paper at hand exhibits the potentialities of the graph-based approach for context sensitive analysis of company-wide data and the direct inclusion of real-time sensor data on-site. The highly dynamic graph-based data model offers the flexibility needed for adapting to changes in the operation of smart, agile manufacturing scenarios.

As shown in the use case scenario the approach ensures that only necessary steps are taken to reach a certain desirable machine status.

The approaches presented in previous publications have been considerably enhanced and the resulting concept presented in this paper has been validated in a laboratory experiment. The lessons learned result in several improvements for the implemented prototype, such as:

- Direct interaction with different source systems other than the IoT platform
- Offering more support to the technician, if need be. The amount and type are to be determined by a survey in the near future
- Adding the capability to visualize the current work status of a certain technician on a remote PC in a web-based client

The future research activity will focus on enhancing the presented approach further than the obvious improvements mentioned above.

To achieve a consistent process that expands into the earlier development phases of products, an import process must be developed to “feed” the graph database with its necessary data. To this extend the import of product structure data is planned, as well as implementing interfaces to synchronize data between the graph-model and a CMMS.

As the task management is already implemented for a single technician, the task management concept will be revised to support on-site collaboration between multiple technicians working on the same facility on large-scale repair or maintenance jobs. The collaboration module will track and coordinate all safety-related stages, and issue coordination instructions when reaching synchronization or combination points in the combined workflow of the technician team.

Taking one step further from the augmented virtuality, the idea of tracking and reviewing the technician’s information from afar, the concept will be brought to the final stage in the reality continuum (cf. Fig. 22.1, [21]), virtual reality. As the prototypical implementation is already able to display 3D visualization of the indications and the concerned asset, only the immersion is left to be sorted out to realize a virtual reality counterpart of the real work for a remote technician to review.

In preparation for the next evolutionary step of the presented concept, we will adapt to a more complex demonstrator machine that features pneumatics, electronics, PLC control, RFID-based sensors and is usually used in professional education of technical personnel.

References

1. Aleksy M, Rissanen MJ, Maczey S et al (2011) Wearable computing in industrial service applications. Procedia Comput Sci 5:394–400. doi:10.1016/j.procs.2011.07.051
2. van Krevelen DWF, Poelman R (2010) A survey of augmented reality technologies, applications and limitations. Int J Virtual Real 9(2):1–20
3. Dini G, Mura MD (2015) Application of augmented reality techniques in through-life engineering services. Procedia CIRP 38:14–23. doi:10.1016/j.procir.2015.07.044
4. Schlick CM, Jeske T, Mütze-Niewöhner S (2014) Unterstützung von Lernprozessen bei Montageaufgaben. In: Schlick CM, Moser K, Schenk M (eds) Flexible Produktionskapazität innovativ managen: Handlungsempfehlungen für die flexible Gestaltung von Produktionssystemen in kleinen und mittleren Unternehmen. Springer Vieweg, Berlin
5. Abramovici M, Göbel JC, Neges M (2015) Smart engineering as enabler for the 4th industrial revolution. In: Fathi M (ed) Integrated systems: innovations and applications. Springer International Publishing, Cham, pp 163–170
6. Reinhart G, Patron C (2003) Integrating augmented reality in the assembly domain—fundamentals, benefits and applications. CIRP Ann Manuf Technol 52(1):5–8. doi:10.1016/S0007-8506(07)60517-4
7. Eisenhauer M, Oppermann R, Prinz W (2007) Internet der Dinge: Anwendung von RFID- und Tracking-Technologien zur intelligenten kooperativen Assistenz im Arbeitsprozess. In: Internet der Dinge. Springer, Berlin, pp 49–62. www.internet-der-dinge.de

8. Wohlgemuth W, Friedrich W (2007) Advanced augmented reality technologies for industrial service applications: ARTESAS; Sachbericht Gesamtvorhaben; zusammenfassender Schlussbericht; BMBF-Leitprojekt ARTESAS; Laufzeit/Berichtszeitraum: 01.02.2004–30.06.2006; Berichtstermin: 30.09.2006, Version: 30.01.2007. Siemens AG, Nürnberg
9. Barthelmey A, Störkle D, Kuhlenkötter B et al (2014) Cyber physical systems for life cycle continuous technical documentation of manufacturing facilities. Procedia CIRP 17:207–211. doi:10.1016/j.procir.2014.01.050
10. Hou L, Wang X, Truijens M (2013) Using augmented reality to facilitate piping assembly: an experiment-based evaluation. J Comput Civ Eng 29(1):5014007. doi:10.1061/(ASCE)CP.1943-5487.0000344
11. Azuma RT (1997) A survey of augmented reality. Presence: Teleoper Virtual Environ 6 (4):355–385. doi:10.1162/pres.1997.6.4.355
12. Abramovici M, Krebs A, Wolf M (2014) Approach to ubiquitous support for maintenance and repair jobs utilizing smart devices. In: Horváth I (ed) Tools and methods of competitive engineering: digital proceedings of the tenth international symposium on tools and methods of competitive engineering—TMCE 2014, May 19–23, Budapest, Hungary. Faculty of Industrial Design Engineering Delft University of Technology, Delft, pp 83–93
13. Neges M, Wolf M, Abramovici M (2015) Secure access augmented reality solution for mobile maintenance support utilizing condition-oriented work instructions. Procedia CIRP 38:58–62. doi:10.1016/j.procir.2015.08.036
14. Partynski D, Koo SG Integration of smart sensor networks into Internet of Things: challenges and applications. In: 2013 IEEE international conference on green computing and communications (GreenCom) and IEEE Internet of Things (iThings) and IEEE cyber, physical and social computing (CPSCom), pp 1162–1167
15. Finkenzeller K, Gebhart M (2015) RFID-Handbuch: Grundlagen und praktische Anwendungen von Transpondern, kontaktlosen Chipkarten und NFC, 7, aktualisierte und erw. Aufl. Hanser, München
16. Schulz D, Gitzel R Seamless maintenance—integration of FDI device management and CMMS. In: 2013 IEEE 18th conference on emerging technologies and factory automation (ETFA), pp 1–7
17. Sean W, Levi L, Steven F (2007) Visual hints for tangible gestures in augmented reality. In: 6th IEEE and ACM international symposium on mixed and augmented reality, 2007, ISMAR 2007, pp 47–50
18. Abramovici M, Krebs A, Lindner A (2013) Exploiting service data of similar product items for the development of improved product generations by using smart input devices. In: Abramovici M, Stark R (eds) Smart Product Engineering. Springer, Berlin, pp 357–366
19. Andrea Z, Nicola B, Angelo C et al (2014) Internet of things for smart cities. IEEE Internet Things J 1(1):22–32. doi:10.1109/JIOT.2014.2306328
20. Dörner R, Broll W, Grimm P et al (eds) (2013) Virtual und augmented reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität. eXamen.press. Springer Vieweg, Berlin
21. Milgram P, Takemura H, Utsumi A et al (1995) Augmented reality: a class of displays on the reality-virtuality continuum. In: Das H (ed) Photonics for industrial applications. SPIE, pp 282–292
22. Haberfellner R, Daenzer WF (eds) (2002) Systems engineering: Methodik und Praxis, 11., durchges. Aufl. Verl. Industrielle Organisation, Zürich
23. Harrison G (2015) Next generation databases: NoSQL, NewSQL, and Big Data: what every professional needs to know about the future of databases in a world of NoSQL and Big Data. Apress (IOUG), New York, The expert's voice in Oracle
24. Vajna S, Bley H, Hehenberger P et al (2009) CAx für Ingenieure: Eine praxisbezogene Einführung, 2, völlig neu, bearb edn. Springer, Berlin
25. Llinas J, Hall DL (1998) An introduction to multi-sensor data fusion. In: ISCAS '98 1998 IEEE international symposium on circuits and systems, pp 537–540

26. Tönnis M (2010) Augmented reality: Einblicke in die Erweiterte Realität. Informatik im Fokus, vol 0. Springer, Berlin
27. Koch C, Neges M, König M et al (2014) Natural markers for augmented reality-based indoor navigation and facility maintenance. Autom Constr 48:18–30. doi:10.1016/j.autcon.2014.08.009
28. Rosen R, von Wichert G, Lo G et al (2015) About the importance of autonomy and digital twins for the future of manufacturing. IFAC-PapersOnLine 48(3):567–572. doi:10.1016/j.ifacol.2015.06.141

Chapter 23
Remotely Piloted Aerial Systems Support Considerations

Jonathan Pelham

Abstract Remotely piloted aircraft (RPA) also known as drones have in recent years become an essential tool for civilian and military users in finding out information about their environment in a cost effective way. Their uses at small scales have encompassed fields as diverse as crop monitoring through to safety critical inspection on oil rigs. Larger RPA have also found a wide variety of niches from scientific research to shipping monitoring and persistent surveillance of military targets. As these systems have seen wider use across a variety of users at various scales, and with a wide variety of uses a picture has started to emerge of how their unmanned nature offers unique challenges to their through life support and to the frameworks that can be used to assure safe operation through life.

23.1 RPAS (Remotely Piloted Aerial Systems)

Remotely piloted aircraft are aircraft that are intended to fly without a pilot on board. They consist of several main system elements. This section will introduce RPAS (Remotely piloted aerial systems) and discuss their use. Some of the difficulties of operating RAS and problems unique to RPAS will be discussed. A generic RPAS architecture is shown in Fig. 23.1. The individual elements of the RPAS shown in the figure are discussed within this chapter and the contribution of each to the through life health management of the system discussed. Briefly the chief portions of the system are the aircraft, the communication means, the GCS (Ground Control Station), the pilot or pilots, and the wider supporting system.

There are many different types of airframe used for the remote aircraft part of the RPAS. Some have become highly recognisable such as the MQ-1 Predator and some have become ubiquitous such as the increasingly popular quadcopters and octocopters. Each type of RPA has different missions to which they are suitable and

---

J. Pelham (✉)
Research/Teaching Fellow in Flight Data, Safety and Accident Investigation Centre,
Cranfield University, Cranfield, England, UK
e-mail: j.g.pelham@cranfield.ac.uk

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_23

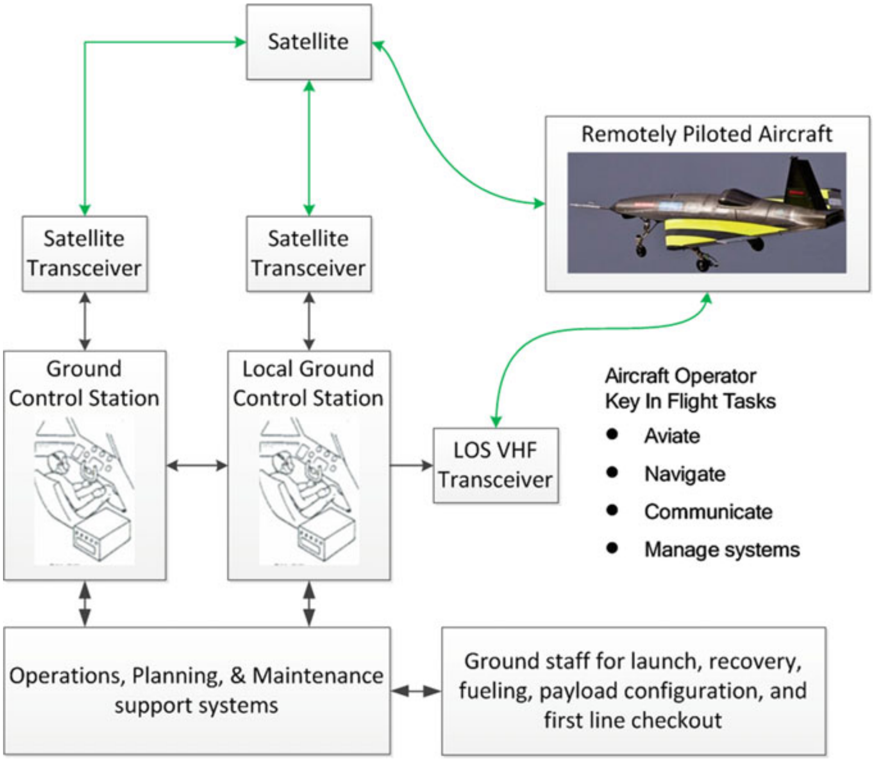


Fig. 23.1 RPAS architecture

different operational challenges caused by the design decisions made during their creation. Quadcopters are highly agile and able to get very close to assets for inspection but have limited payload capacity and endurance due to the high power requirements of hovering flight. Long endurance aircraft such as Predator are able to stay aloft for multiple hours and take very detailed pictures and collect other useful sensor data but are highly complex and require sophisticated support structures to be in place to ensure they are used effectively.

The ground control station is a long winded way of describing the way by which the remote pilot controls the aircraft. This can be done at various levels of control all the way from full control authority, through high level mission planning, to just supervisory control of the aircraft. Whatever level of control chosen the remote pilot remains responsible for the safety of the aircraft and the safe interaction of the aircraft with other airspace users and potential 3rd parties on the ground. The choice of a ground station is often rather limited as it is frequently supplied as part of the control system for the aircraft selected for use. There are however emerging standards for GCS and for communications between the GCS and the RPA. Particular attention should be paid to the Dronecode project which is creating a full stack of open source GCS software, aircraft flight management, and communications systems [1].

The communication system that connects the GCS and the RPA could be one of many different methods. Frequently quadcopters use radio in the 2.4 GHz band which can result in relatively short range. Every method to communicate with the aircraft will have some pros and cons that affect operations and must be understood. The USAF controlled the Predator through a combination of VHF radio for line of sight control during take-off and landing and satellite band radio for control during the cruise and mission phases of the flight. Control via satellite relay created large amounts of latency which reduced the ability of the remote pilots to understand potential aircraft issues. It did however allow control from almost anywhere in the world.

The operation supporting means of the RPAS is the entire team that manages, repairs, supplies, and supports the RPA. The team on the ground that prepares the aircraft for transport between mission locations are every bit as responsible for flight safety as the pilot despite a certain detachment from the immediacies of flight control.

23.2 RPAS Mishaps and Challenges

RPAS experience challenges in just the same way as manned aircraft but with some additional complicating factors. The average DoD UAS mishap rate has been claimed to be as high as 50 per 100,000 flying hours (DoD manned mishap rate claimed to be 1 per 100,000 flying hours) [2]. Figure 23.2 shows the mishap rates for the USAF (United States Air Force) MQ-1 Predator, RQ-4 Global Hawk, MQ-9 Reaper, U-2 Dragon Lady, F-16 Falcon, and F-22 Raptor. The U-2, F-16, and F-22 are all manned platforms and it can be observed how with increasing fleet hours the mishap rate of these aircraft decreases significantly. The initial conclusion could be that this shows broadly comparable mishap rates between manned and unmanned aircraft. However the F-16 and F-22 are both supersonic fighter aircraft and the U-2 has a notorious history of being a very difficult aircraft to fly. To put this in context the Boeing 737 has a class A mishap rate per hundred thousand flying hours of around 0.13 and that performance is more than an order of magnitude better than the best performance in the figure. If RPAS are to live up to the many potential opportunities for their use then their mishap rate will need to be substantially improved. There is very little data available for smaller platforms used by civil operators. There is anecdotal evidence that the mishap rate for small platforms is higher than that shown in Fig. 23.2. However without hard evidence it is difficult to come to further conclusions.

As of 2009 the USAF had accrued a total of 675,450 flying hours with the Predator, Global Hawk and Reaper. A study by Hartfield et al. [3] published in 2012 went through the mishap data for these aircraft. The mishaps were classified

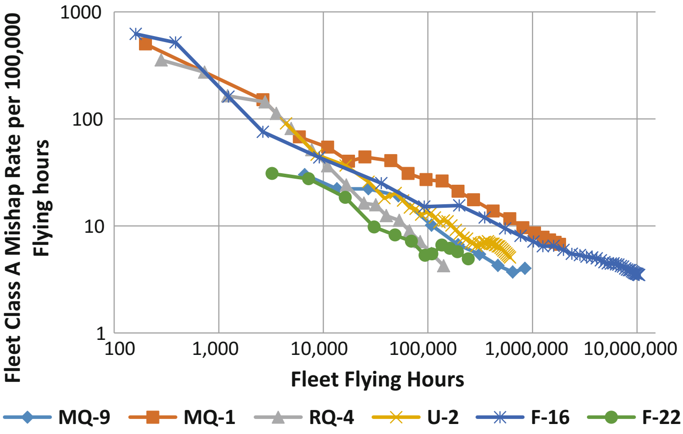


Fig. 23.2 USAF RPAS mishap rates

by the USAF [4] and the study further examined their data to identify mishap cause and system.

The total accrued mishap rate was 57.3 mishaps (A–E) per 100,000 h. The rates for Global Hawk and Predator B/Reaper are projected as during the period of the study the respective fleets did not exceed 100,000 flying hours.

Individual mishap rates (classes A–E, all categories) [3]

Predator	33.7 Mishaps per 100,000 h
Predator B/Reaper	64 Mishaps per 100,000 h
Global Hawk	53.8 Mishaps per 100,000 h

Accrued class A mishap rates (all categories) [3]

Predator	11 per 100,000 flying hours
Predator B/Reaper	18.8 per 100,000 flying hours
Global Hawk	12.7 per 100,000 flying hours

These three aircraft all share a number of common features as well as being unmanned. They all feature high aspect ratio wings, a single engine, satellite modem for beyond line of sight control, and long endurance. These aircraft cannot currently refuel in flight and thus this long endurance suggests a large take-off weight fuel fraction. Over the 10 years of this study (2001–2009) there were 79 Class A mishaps but only 11 Class B mishaps ($200,000 to $999,999 in Damages). There are also 61 class C mishaps over the period in question. This could suggest RPAS are more prone to complete write off or low damage than a manned aircraft.

It was suggested by Hartfield that these RPAS may be susceptible to single item catastrophic failure due to their design [3]. The largest single mishap cause identified was reliability at approx. 57% of class A mishaps [3]. This was dominated by power and propulsion reliability which should perhaps be expected as these are all single engine aircraft. It certainly shows the need for careful maintenance of RPAS.

The U-2 manned reconnaissance aircraft shares some of the features of these unmanned aircraft and also has a low rate of class B mishaps compared to its class A mishap rate [5]. The U-2 similar features include a high aspect ratio wing and a large fraction of its take-off weight is fuel. The very low wing loading during landing makes these aircraft increasingly susceptible to gusts as fuel is burned and contributes to difficulty during the final touchdown phase due to ground effect. RPA suffer more from this as their wing loading can be as low as a third of that expected for a manned aircraft [6]. The study revealed that these aircraft suffered 43 mishaps categorized as Heavy Landing during the period under study of which 15 were Class A mishaps. This demonstrates that the aircraft handling can be compromised by mission optimisation. RPAS operators would do well to consider that as well as the challenges due to remote operation they face the additional burden of aircraft with a very narrow margin of safety engineered to fly for as long as possible.

One of the most widely discussed RPAS mishaps was the crash of a US border patrol Predator near Nogales, Arizona [7]. In this incident the GCs console operating system experienced a crash and the remote pilot switched to use the camera operators console but did not follow console handover checklist. This meant they failed to check the position of the control levers and when the switchover was completed the lever positions caused the Predator engine fuel flow to be shut off during flight. The engine was starved of fuel and the aircraft crashed. This turned a slow problem with a crashed pc console into a quickly deteriorating issue with a crashing aircraft. It shows the core need for checklists when performing complex activities and also the need to avoid design decisions which can enable simple mistakes. Designing systems to be less brittle in their tolerance of unintended actions is a field of discussion out of scope to this chapter but worthy of further study.

Manning et al. [8] blames ground control station design for ongoing problems in the mishap rate of unmanned aircraft. This complexity of control is coupled with the checklists that sometimes get ignored to the detriment of safe operation. Not following checklists was identified as a contributory factor in several mishaps [9–12].

The key problems identified in Predator mishaps [13, 14]

- Skills and Knowledge
  - Checklist error
  - Task miss-prioritization
  - Lack of training for task attempted
  - Inadequate system knowledge

- Situational Awareness
  - Channelized attention
- Crew coordination

These are all relevant to operators of other RPAS whether military or civil operators.

RPAS Maintenance

A review of the studies into RPAS maintenance shows one of the most common recommendations found is the keeping of proper logs to record maintenance tasks performed on the aircraft or supporting systems. This ties in well with the more general need to document procedures and to make checklists to more conveniently enable staff to actually follow those procedures. See The Checklist Manifesto for further discussion [15]. Even the big ticket aircraft are sometimes developed in a rush. Figure 23.2 USAF RPAS Mishap Rates shows how the mishap rate of the MQ-1 was very high initially but has improved over time as the USAF has learned how to operate the Predator and the Predator has been improved based on lessons learned. One crucial part of this improvement was an improvement in the maintenance procedures and the maintainability of the Predator system. Prior to the Nogales Predator crash cited earlier there were several incidents of the GCS locking up but no maintenance action was taken to address this issue. The maintenance of the GCS and associated equipment is every bit as necessary as proper maintenance of the aircraft.

23.3 RPAS Through Life Support

The support of the RPAS through life consists of a process going from acquisition where the specifications for the required RPAS are determined, through operation, to disposal. The through life management of assets is discussed well elsewhere and here the focus will be on the management of the RPAS itself.

Figure 23.3 shows an illustration of the decision making process for RPA actions. Its intent is to show the need for data and information to be considered when setting strategy and also that the time available to make the decisions is different for different parts of the system. For example an aircraft may not be able to usefully discuss the relative merits of air safety policy chances but would be able to respond to a gust during flight in a much more timely and appropriate manner than the pilot on the ground. This difference in capability, role, and the need for an appropriate level of decision making is a key question for effective operations that to some extent will always be held in tension by the capabilities of operators, systems, and algorithms. The decision about how an operator will operate may be made by adopting a recognized framework and adhering to an operations manual of known provenance. It may also be made by working from first principles to choose

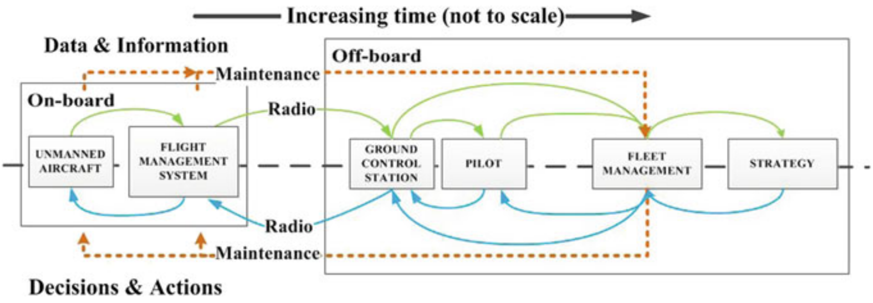


Fig. 23.3 RPAS decision timescales

appropriate processes, aircraft, and support for specific missions. In some circumstances the decision may be an unconscious one but the end result may not be desirable. In a bit more depth the right most box of the figure shows strategy whereby the strategy of the operator is considered, set, and applied to operations. The operator strategy must be informed by data and information from fleet management, from the pilots who conduct operations, and where possible from data developed from actual operations. Data mining can be used to review the flight data retained from previous missions to look for trends that could suggest a need for changes to operational procedures or training for personnel. This is the same in concept to FOQA systems used by large airlines [16, 17].

The strategic decisions taken have wide consequences and if a strategy is not set the operations may not reflect the culture desired or conform to local regulations. The absence of a safety management strategy still represents a choice albeit a negative one. Fleet management is informed by the strategic choices of the company and uses information from staff and systems in the day to day management of the fleet. Operational problems are reported and problems rectified in a timely manner. The time available for these decisions is less than those decisions taken at a strategic level and those making them may find themselves more able to understand fleet problems due to their involvement in fleet management. The next two boxes are the GCS and the pilot. These are shown as separate because the pilot will not automatically understand data the GCS is displaying for their attention. To paraphrase Billings et al. [18] data does not become information until it is understood in the mind of the pilot. In addition the GCS could be programmed to make automated mission decisions based on aircraft and other data in a timeframe that the pilot would not be able to match. In this type of decision the pilot would have to be aware of the procedures and able to supervise appropriately. The left most box represents the aircraft and its flight management system. The decisions and actions that can be taken by the aircraft can occur within time constraints that could not be matched by any off board system but will necessarily be limited in the amount of computational power available to make those decisions. Typical decisions made include flight stabilization and aircraft pitch required to achieve level flight at a requested speed.

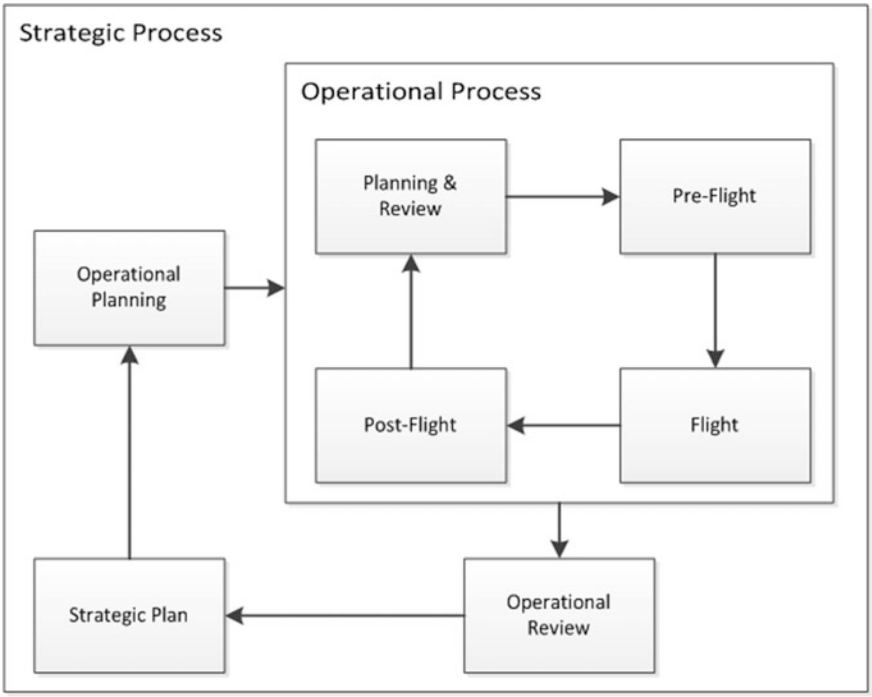


Fig. 23.4 UAS management

Figure 23.4 shows a generic UAS operator process to illustrate the different elements which contribute to the safe operation of the aircraft. To illustrate the application lets go through the typical operation of an aircraft within this scheme. It is agreed to go and do a survey of the land of a client. The operator consults the operational plan and makes any necessary modifications for this particular project. Required tools, parts, aircraft, and team members are accounted for and a plan made to get them all to the work site at the appropriate time. Experienced personnel may be able to do these items much more quickly as they readily understand what is needed from their experience but the use of a plan and a checklist reduces error and helps less experienced personnel contribute as they develop their own skills. When the team arrives on site the plan of work is examined and reviewed based on a site survey and any unexpected risks discussed and assessed. The aircraft is then prepared for flight and its performance checked in accordance with the operations manual of the operator. During flight the pilot and any supporting staff keep a sharp eye for problems and take any appropriate mitigating actions. After flight the aircraft is turned around to prepare it for the next flight and mission data downloaded. The site survey being performed may consist of several flights and it is appropriate between each one to check that conditions are still right for flight and that no new risks have emerged. The process of safe operations is performed through continual review of risks and the application of control measures to mitigate their likelihood or impact. This will be done internally by any experienced pilot but the risks and control measures should be documented to ensure they are retained and learned from instead of trusting to the vagaries of individual memory. After the operation to conduct the survey has been completed and all equipment packed and transported back to base it is appropriate to review and note down any lessons learned or interesting occurrences that can help guide the planning and preparation for future operations. These can be as mundane as discussing whether taking additional batteries would have sped up the completion of the mission to a review of whether guidance on weather restrictions for operations is appropriate or should be tightened. The through life support of RPAS is a process of constant review and decision making. It represent at its core the OODA loop discussed by Boyd [19]. Observing your environment, orienting the observation into context, deciding on your action, and then taking the action. This iterative loop of decision making and evaluation decision results is the constant process of a competitive business. It can also be expressed in a similar way to the evolutionary paradigm. There the expression would be that the fittest survive and in this context we can suggest that the best informed survive to thrive in RPAS operations. It is entirely possible to operate an RPAS in accordance with a manual you have adopted, to have no mishaps, and to have a successful business. If improvements in safety, efficiency, and cost effectiveness are desired however then not only will some thought be required but the also the experience and analysis of mistakes. If lessons can be learned from the mistakes of others then so much the better.

23.4 Conclusion

RPAS are complex even when consisting of a single aircraft and a single pilot. To manage these systems safely and effectively some amount of thought and methodical decision making is required. The use of checklists is an effective way to manage the performance of critical steps necessary for safe operation [15]. In order to consider how safety and usage can be improved through life RPAS operations and fleet management should undergo a consistent process of review and strategies for operations improvement discussed with operators and then set out to be followed. The consideration of how repairs will be conducted, spare parts will be found, who will fit them, and how the effectiveness of repairs will be assessed should be given consideration and plans put in place. The choice of an RPAS that cannot be repaired effectively can cripple the ability of the organisation to complete contracts in a timely and professional manner. The process by which an aircraft is deemed obsolete or unsafe must also be known so that when the aircraft reaches that state it can be removed from active status and either used as a parts source to keep other aircraft operational or responsibly disposed with. In conclusion even the smallest operator has the ability to make a checklist to aid their operations and if they wish to improve their operations and maintain a minimum standard then they would be well advised to do so.

References

1. Dronecode Project (2015). https://www.dronecode.org/. Accessed 21 Sept 2016
2. ("Raza") Waraich QR, Mazzuchi TA, Sarkani S, Rico DF (2013) Minimizing human factors mishaps in unmanned aircraft systems. Ergon Des Q Hum Factors Appl 21(1):25–32
3. Hartfield R, Carpenter M, Randall W, Hundley J (2012) Unmanned air vehicles (UAV): safety event prediction, and classification. In: 12th AIAA aviation technology, integration, and operations (ATIO) conference and 14th AIAA/ISSMO multidisciplinary analysis and optimization conference
4. Redmond W (2008) Air force instruction 91-204: safety investigations and reports. Secretary of the Air Force, Washington
5. USAF Safety Centre (2015) U-2 mishap history
6. Weissberg V, Green A, Mey-Paz H (2012) The birth of specialized airframe technology for UAVs. In: 52nd Israel annual conference on aerospace sciences 2012, vol 1, pp 324–336
7. D. No, E. No, Guide FCMS (2008) National Transportation Safety Board: GAO-08-652T, Washington, USA
8. Manning SD, Rash CE, LeDuc PA, Noback RK, McKeon J (2004) The role of human causal factors in U.S. army unmanned aerial vehicle accidents
9. Hobbs A, Herwitz SR (2005) Human factors in the maintenance of unmanned aircraft. Small 15(100):100s
10. Tamilselvan P, Wang P, Jayaraman R (2012) Health diagnostics with unexampled faulty states using a two-fold classification method. In: 2012 IEEE international conference on prognostics and health management: enhancing safety, efficiency, availability, and effectiveness of systems through PHM technology and application, PHM 2012
11. Whittle R (2011) Predator's big safari. Arlington, Virginia Mitchell Institute. Airpower study, vol 14. http://www.afa.org/Mitchell/Reports/MP7_Predator_0811.pdf
12. Williams KW (2004) A summary of unmanned aircraft accident/incident data: human factors implications. Defense Technology Information Center Document, Ft. Belvoir, VA, 22060, RPRT
13. Nullmeyer R, Herz R, Montijo G (2009) Training interventions to reduce air force predator mishaps. In: 15th international symposium on aviation psychology
14. Montijo GA, Kaiser D, Eberhart J, Spiker VA, Nullmeyer R (2010) Training interventions to reduce predator crew errors. In: The interservice/industry training, simulation and education conference (I/ITSEC), vol 2010, no 1
15. Gawande A (2011) The checklist manifesto: how to get things right
16. Flemming J (2012) Utilizing FOQA (Flight Operations Quality Assurance) in training. In: 57th annual corporate aviation safety seminar 2012, CASS 2012, pp 52–60
17. EAFDM (European Authorities Coordination Group on Flight Data Monitoring) (2013) Developing standardised FDM-based indicators
18. Billings C (1996) Human-centered aviation automation: principles and guidelines. NASA, Ames Research Center, Moffett Field, CA 94035 United States, RPRT
19. Osinga FPB (2007) Science, strategy and war: the strategic theory of John Boyd. Routledge

Chapter 24
Preventive Maintenance Scheduling Optimization: A Review of Applications for Power Plants

Neha Prajapat, Ashutosh Tiwari, Xiao-Peng Gan, Nadir Z. Ince and Windo Hutabarat

Abstract This review paper considers literature in the field of preventive maintenance scheduling optimization, particularly for applications to power plants, with the view to assess the methods used and identify research trends and gaps. Aspects of each of the papers such as application domain, problem formulation, model formulation and optimization techniques have been analyzed and assessed. Research trends such as the increasing use of stochastic parameters, multiple objectives and hybrid optimization methods have been identified. A research gap has been identified: the application of discrete-event simulation methods with multi-objective hybrid optimization for power plant preventive maintenance scheduling. These areas provide exciting research opportunities with significant potential benefits for power generation companies including increased profit and reliability.

24.1 Introduction

Various different types of maintenance scheduling can be defined: Corrective Maintenance (CM), Preventive Maintenance (PM), and Condition Based Maintenance (CBM). CM is performed upon component failure and generally leads to high failure cost and large downtimes [1]. Monitoring technologies are often used to perform on-line assessment of the condition of equipment and then condition based maintenance activities can be applied to prevent high failure costs. Prognostics can be used as an extension of condition based maintenance strategies. Prognostics use condition based maintenance technologies to diagnose failures of components and to potentially forecast future failures [2]. Preventive maintenance

---

A. Tiwari (✉) · W. Hutabarat
EPSRC Centre for Innovative Manufacturing in Through-Life Engineering Services,
Cranfield University, Cranfield, UK
e-mail: a.tiwari@cranfield.ac.uk

N. Prajapat · X.-P. Gan · N.Z. Ince
GE Power, Newbold Road, Rugby, UK

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_24

can be used in conjunction with, or as an alternative to, monitoring based maintenance strategies. Preventive maintenance constitutes a predefined schedule or routine of check-ups and repair tasks [3]. The focus of this review is on optimizing preventive maintenance schedules.

Preventive maintenance scheduling has been widely used within industry and these activities are often performed at a higher frequency than is required to prevent unnecessary failures [4]. The manufacturer or equipment supplier will usually predefine a set of constraints and maintenance intervals for the schedule to abide [5]; these are often based on the experience of engineers. A number of industry sectors have identified the benefits which can be achieved through optimization of preventive maintenance schedules, for instance: increased reliability and profit and lower risk and costs. Applications for preventive maintenance scheduling optimization have appeared in a range of industry sectors including transport [6, 7], manufacturing [8] and power industries [5, 9, 10]. This work is concerned with the application of preventive maintenance scheduling optimization for power plants.

Optimization of maintenance schedules within the power generation industry is vital as generation equipment must be reliable, to enable generation companies to provide a reliable service to customers. This optimization could also produce large cost savings through increased sales and decreased downtime, also increasing the overall utilization of the plant. These savings made in operational costs can also assist generation companies by enabling them to provide competitive energy prices [11]. For larger power plants the main equipment can have different manufacturer's guidelines and maintenance intervals; in such a scenario it could be beneficial to perform simultaneous maintenance on major equipment as the opportunity arises.

A number of review papers considering maintenance scheduling optimization exist: in [12] a detailed literature review is performed of scheduling research covering job shop scheduling, work in [13] focusses on multi-objective production scheduling research. These reviews cover the generic machine scheduling problem and do not consider preventive maintenance scheduling within the power industry. In [1] authors provide a brief overview of different levels of maintenance scheduling within power systems and outline the maintenance scheduling problem in power systems for various time scales. Studies [3, 14] examine literature for the Generator Maintenance Scheduling problem (GMS). In [14] authors provide details on problem features, goals for optimization and optimization techniques. Authors in [3] provide a comprehensive description of problem features, objective functions and interfaces with other scheduling problems. This paper presents an up to date review of papers within all areas of power plant preventive maintenance scheduling optimization with detailed information about objective functions, constraints and optimization methods. In particular this paper examines various optimization methods for the preventive maintenance scheduling problem including: dimensionality reduction [4], expert systems [15], Genetic Algorithms (GA) [9, 16–18], hybrid methods [19–21], Particle Swarm Optimization (PSO) [22–24], Simulated Annealing (SA) [25] and Tabu Search (TS) [26]. Notably the application of hybrid optimization methods is a new research area which has not been covered by previous reviews.

24.2 Methodology

Scopus, being the largest abstract and citation database for peer-reviewed literature, was chosen as the main database to access papers. Initially scheduling optimization review papers were studied to provide a good overview of scheduling optimization within industry. Subsequently a number of review papers for maintenance scheduling were found and used to identify journals and conferences which contain the most relevant papers. Papers were chosen based on relevance to preventive maintenance scheduling with application to the power industry and optimization.

A primary search in Scopus using the keywords: ‘maintenance scheduling power’ resulted in over 1000 results. Hence, a more focused search was made using the keywords: ‘preventive maintenance scheduling optimization power plant.’ This search returned around 25 papers among which 11 papers were chosen. The number of papers which focus on power plant equipment maintenance is limited; hence papers which explore the GMS problem were found through a further search using ‘generator maintenance scheduling’, as generator has been used as a synonym for power plant by many authors. There is a large amount of literature on the GMS problem, hence a subset of GMS papers were chosen. Papers were chosen which specifically applied optimization methods and also had more citations. This paper focusses on the optimization methods applied by papers in the area of preventive maintenance scheduling for power plants.

Information about 45 chosen papers was collated in an Excel spreadsheet. Columns within the spreadsheet include: year of paper, journal or conference, country, name of paper, authors, objective functions, constraints, application area, optimization method and any further comments. From the spreadsheet, trends were easily identified using graphs. For instance, the column titled ‘year’ was analysed and a trend indicating an increasing amount of research within recent years was identified; Fig. 24.1 illustrates this trend.

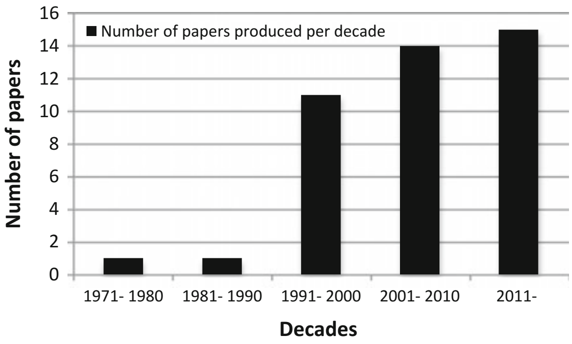


Fig. 24.1 A graph to show the number of papers produced per year group

24.3 Application Domains

The optimization of preventive maintenance schedules has been applied on various different levels within the power sector. Preventive maintenance scheduling has been applied to a range of types of plant: nuclear [27], gas [5], hydro-electric [28] and wind [29] among others. In addition to generation, distribution services have had their maintenance schedules optimized. For example, transmission maintenance has been optimized [30, 31]. The majority of papers consider the Generator Maintenance Scheduling problem, which typically involves maintenance on a number of equipment types across a number of power plants. Only seven of the papers studied addressed preventive maintenance scheduling for single power plants.

24.3.1 Multiple Generator Maintenance Scheduling

The optimal maintenance scheduling of multiple generators is imperative as generators govern the routines of other equipment and particularly because a number of planning activities are based around this schedule [9]. The maintenance scheduling of systems of power plants has been described as a 'challenging optimization problem,' [14]; this is mainly due to the increasing complexity of models within recent years and the enormous scale of the problem. In particular, the reliability of the power supply for customers is crucial and hence this problem has received a lot of academic interest since the first paper was produced.

Many very large scale power maintenance scheduling problems have been proposed. In [28, 32, 33] the authors consider large scale systems with around 80 units including thermal, hydroelectric and nuclear plants. The author extends previous work in [29] to include maintenance scheduling of wind and hydroelectric units, which are becoming increasingly important within the power generation industry. Another large system is provided in [15]. This study looks at the scheduling of 33 generation units and 179 transmission lines.

The IEEE Reliability Test System (RTS) has been widely used [10, 34, 35]. Work in [16] modifies and extends the IEEE RTS to include more units and 38 transmission maintenance lines.

The vast majority of papers consider the maintenance of between 20 and 35 units [18, 30, 36, 37]. Wind and nuclear units are also considered in paper [18]. A comparison of results for various test cases is made by some papers [34, 38]. A much smaller system has been considered by [27], the system has only 4 units.

Studies have also considered the optimization of maintenance schedules for various plant components. In [26] authors assume that the generating unit schedule is fixed and they align the outages for 61 other maintenance tasks. In [11] authors optimize a smaller scale problem for 42 pieces of equipment in total within 2 co-generation plants.

24.3.2 Single Power Plant Maintenance Scheduling

The problem of optimizing the maintenance schedules of components within a single power plant has also been studied within literature. The study in [5] considers the gas turbine preventive maintenance scheduling problem; the gas turbine is considered to be a deteriorating system. Authors in [23] consider the maintenance schedules of the components of a nuclear power plant separately. These components must be maintained and all possible scheduling combinations must be considered. The study in [39] looks at single nuclear plant subsystem availability and similarly [40] optimizes the maintenance scheduling of nuclear power plant sub systems. Work in [41] looks at tasks based in maintenance classes for subsystems of a coal fired plant. Authors in [4] collect data from 18 turbines to determine ideal maintenance intervals.

24.4 Problem Formulations

24.4.1 Objective Functions

Within the field of preventive maintenance scheduling, the vast majority of papers acknowledge the importance of cost and reliability for the MS optimization problem. These criteria are conflicting and this is particularly significant when the system is maintained by an operator who is independent from the generation company. Amongst the papers which were analyzed for this review, all of the papers used cost or reliability or some combination as the basis of their objective functions for optimization. Figure 24.2 indicates the proportion of

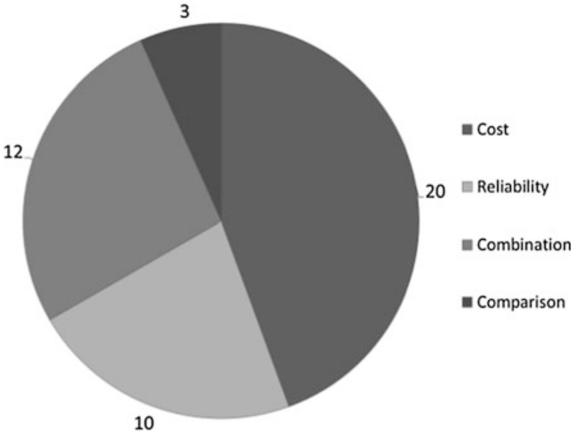


Fig. 24.2 A pie chart to show the objective functions of papers studied

papers which use cost, reliability, both objectives and which compare each criterion separately through test cases. From this pie chart it can be seen that cost is used as a primary objective function by almost half of the papers, which is the largest proportion.

24.4.1.1 Cost Minimization

Numerous studies consider cost minimization as the main objective for optimization [27, 30, 31, 33, 38, 41–46]. Cost based objective functions fall under either profit based or cost based objective functions. The cost based objective functions often account for the cost of maintenance and the cost of production or operations [47], whereas profit based objective functions take into consideration factors such as fuel cost and electricity price.

A cost approach with the aim to minimize the total maintenance and production costs has been considered within studies [4, 30, 31, 44, 48]. In addition to these costs, [33] also considers the start-up costs within the objective function. In [45] authors append transmission maintenance costs to generator maintenance costs and operational costs. In [17] minimization of expected energy production cost is used and maintenance costs are disregarded. On the other hand, [15, 41] consider the minimization of maintenance costs to be their primary objective. In [49] a detailed profit based approach is applied. A stochastic optimization method is applied to deal with the uncertainties of electricity and fuel prices. Work in [24] uses the minimization of Maintenance Investment Loss (MIL) as a profit criterion. Authors in [43] set the sale of electricity according to market clearing price forecast, this is used as an objective for maximization.

A combination of both profit and costs are considered in a number of studies. For instance [27] consider the ‘spread,’ (the difference between energy price and the cost of generation) for UK plants. Profit and cost based objective functions are modelled individually and then compared in [5] for the sequential problem formulation. The study concludes that the profit based formulation could potentially improve profitability for the plant.

Other cost based objectives are proposed in a number of studies. For instance utilization maximization in [11] could be considered as a cost based optimization criteria as it effectively increases profitability. Study [41] also increases availability of plant and reduces cost through optimization. Another cost based method proposed by [38] is the ‘minimum possible disruption of an existing schedule’. In this particular example an optimal schedule is devised and changes must be incorporated in the schedule in a way that minimizes the overall cost.

24.4.1.2 Reliability Maximization

Reliability maximization is considered to be a highly significant objective function for a number of different studies, particularly for the GMS problem. This objective is considered by many papers [9, 18, 25, 34, 36, 37].

The GMS problem is to serve to a number of generators whilst ensuring that the customer demand for electricity is met. The main objective function used for the GMS problem is to level the reserve rate; the difference between the total capacity of the available units and the demand [9]. If the demand is not met at any point in time, this could result in power outages for customers and a poor reputation for the provider. Thus the reliability of the supply is critical. Papers considering the GMS problem account for levelling the reserve rate in some form and many consider it to be a primary objective function [9, 17, 18, 21, 25, 36, 42].

Another form of reliability criteria for the GMS problem is the Loss of Load Probability (LOLP); this has been applied by a number of papers [14, 20, 34, 42, 47]. This is defined as the likelihood that the system demand for electricity will surpass the available capacity; this can be used to evaluate the risk for each individual time period [34].

24.4.1.3 Combination Approaches

Approaches which combine a number of objective functions have been proposed within literature such as weighting methods and vectors of objective functions. A weighted sum method is used in two studies to combine objective functions and coefficients [19, 47]. For example, [19] use the weighted method to minimize the fuel and maintenance costs and to level the reserve rate. More than two objective functions have also been proposed by a number of studies. In [50] authors simultaneously consider three objective functions: reliability maximization, fuel cost minimization and constraint violation minimization. Similarly [16] use a triple objective function consisting of reliability, risk and economic objectives. A non-dominated solution set is created, hence it is concluded that the method is preferable to a weighting method. In [24] the competing objectives for the servicing and the generation company are both modelled as objectives.

24.4.1.4 Comparison Approaches

A number of studies using comparisons of cost and reliability criteria have also been proposed. For instance cost and reliability have been considered separately as objective functions and the results have been compared by [17, 47]. In [47] production costs and reliability are compared and it is concluded that reliability is a better objective function, as maximizing reliability simultaneously minimizes the production costs. In [17] authors find that their proposed method performs well for integer GAs irrespective of which objective function is applied.

24.4.2 Constraints

It is suggested in [14] that constraints can be categorized into three main types: power generation constraints, resource constraints and technological constraints. Power generation constraints ensure that customer demand is met, resource constraints include manpower and inventory limitations and technological constraints are applied by manufacturers on maintenance intervals. Figure 24.3 shows the number of papers which apply each of the constraints. The constraints are split into the sections by type.

Studies considering a number of constraints have shown that the application of constraints often reduces the objective function values. Authors in [49] demonstrate that the risk constraint reduces the profit function. In [42] integer programming is applied to a test case with exclusion, sequence and reserve constraints for two objective functions; the impact of constraints on the objectives is demonstrated. Authors in [30] consider two different case studies using reliability and transmission security as different constraints, they show that there are significant impacts on the costs when transmission limits are applied to the problem.

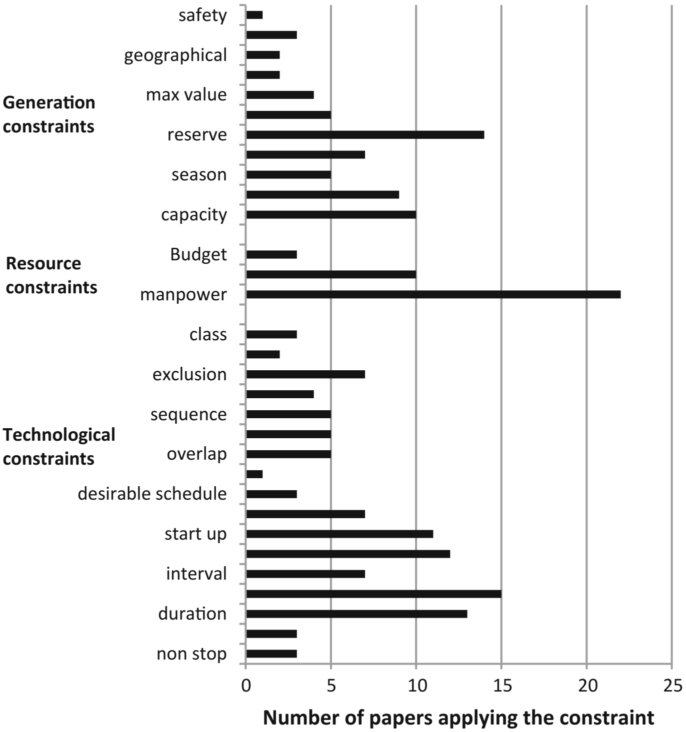


Fig. 24.3 A bar chart to show the number of papers applying each constraint

Studies have fuzzifed the constraints so that the constraints are soft in order to evaluate the trade-off between the constraints and the objective functions. For example, authors in [21] fuzzify the manpower, maintenance window and geographical constraints. Similarly authors in [9] use a fuzzy evaluation function to consider the flexibility of the manpower constraint and compare to alternative methods. The fuzzy evaluation function is able to effectively deal with the trade-off between reliability and manpower within the allowed flexibility.

A number of papers compare two objectives separately, setting one as a constraint and the other as an objective function [21, 40, 42, 47]. Some studies use cost as the objective function and then formulate the reliability as a constraint, these are under power generation constraints. The manpower constraint is the most common constraint among all papers studied. Authors in [9] claim that requiring up to 5% extra manpower might be allowable for a generation company as extra manpower can be hired if this produces better solutions with respect to the objective functions. Extra manpower can be costly for companies; hence the manpower constraint is a cost based constraint. It is clear from the objective functions pie chart that cost is a very significant factor for the optimization and this explains the large number of papers considering manpower as a constraint.

Maintenance duration is defined as a technological constraint [14] and hence this is a highly important constraint which cannot be overlooked by studies, hence it is also considered by a large number of papers. Demand and supply is a power generation constraint used to ensure that the power production is at least equal to the electricity demand [33]; hence this is a reliability based constraint and is applied by a number of papers.

24.5 Model Formulations

Almost all papers studied in this review formulate the maintenance scheduling problem as a mathematical programming problem and then proceed to solve this problem. The most common formulation is integer programming formulation which has been widely used [9, 11, 19, 26, 30, 31, 33, 34, 38, 42, 49–51]. Other formulations include dynamic programming [21] and discrete programming [18, 27] among others. The first part of this section addresses how the problem has been formulated mathematically and the second part addresses stochastic formulation of evaluation functions which is an area of growing interest.

24.5.1 Schedule Formulations

Mathematical programming formulations such as integer programming, mixed integer programming and dynamic programming and decomposition methods have been mainly used to formulate models.

24.5.1.1 Integer Programming Models

Integer programming and mixed integer programming formulations have been proposed by a significant number of studies with several advantages. In [50] integer programming is suggested to be a convenient approach to the resource allocation problem; the main shortcoming is suggested to be the limited accuracy of power system simulation as uncertainties cannot be modelled [47], in addition the method exceeds computational limits. Conversely authors in [42] suggest that integer programming approach is the only true optimization approach which is practical for the problem.

Dopazo in [38] presented the first 0/1 integer programming representation for the GMS problem. For each period and maintenance unit a binary integer is assigned. One of the weaknesses of the approach is identified to be that only one constraint can be applied at a time [30]; this approach has been applied by a number of studies [11, 19, 31, 35, 42].

An alternative to the 0/1 integer programming representation is presented by [9]; in this approach integer variables can be used to represent the period in which maintenance of each unit starts. Authors in [17] also apply this approach to find the starting periods for each maintenance activity.

24.5.1.2 Dynamic Programming Models

In [47] authors claims that dynamic programming is best suited to the optimal preventive maintenance scheduling problem; suggested due to the sequential nature of the problem. In [14] it is agreed that dynamic programming can only be applied to a problem which has previously been formulated as a sequential decision process. In [21] authors formulate the problem as a dynamic programming problem with fuzzy objective functions.

24.5.1.3 Decomposition Approaches

Bender's decomposition method has been applied to maintenance scheduling problems by splitting the original problem into a master and a sub problem [33]. This method can be used after a problem is formulated as an integer programming problem. Authors in [17] claim that a disadvantage of the decomposition approach is that it cannot accurately simulate power system operation due to the curse of dimensionality. In [31] Bender's decomposition is applied to the preventive maintenance scheduling problem. This approach is extended by authors in [33] to also include the transmission and network constraints.

24.5.2 Stochastic Formulations

The application of stochastic methods to evaluate solutions, as an alternative to deterministic methods, within maintenance scheduling optimization has grown within recent years. Studies have used these methods to deal with the uncertainty involved in reliability and costs for preventive maintenance scheduling. These uncertainties of the preventive maintenance scheduling problem for power plants render the optimization problem more complex.

In [35] the Weibull distribution is used to model the failure rate due to component deterioration and the exponential distribution is used to model the random failure rate. Similarly authors in [20] introduce a stochastic objective function to consider daily load variation and random outages. In [24] unavailability of the system is modelled using the Weibull distribution. Within [21] the membership functions of both objectives are fuzzified to deal with the uncertainty in the maintenance scheduling problem. Monte Carlo Simulation has been considered by [49]; it used to model the uncertainty for the price of fuel energy. The results show that for particular units, the use of stochastic prices considerably reduces the maintenance hours.

The use of methods such as discrete event simulation to deal with dynamic uncertainties has arisen in other fields for preventive maintenance scheduling. This could also be applied to the preventive maintenance scheduling problem for power plants to explicitly deal with the uncertainties in prices, reliability, load variation and random outages. These stochastic methods will become increasingly important with the uncertainty related to renewable energy sources.

24.6 Optimization Techniques

A range of optimization methods have been explored within studies to optimize the maintenance schedules within the power industry. The pie chart in Fig. 24.4 shows the different optimization methods used to solve the maintenance scheduling

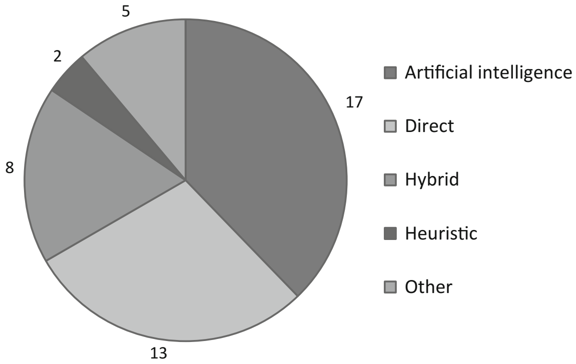


Fig. 24.4 A pie chart to show optimization techniques used

problem. Artificial intelligence methods are the most common method; Genetic Algorithms were used in 7 studies. Direct methods have also been applied to solve mathematical programming formulations. Hybrid methods have been used by 7 of the studies, to combine artificial intelligence, heuristic and direct methods; this is an area of growing research. Other methods include dimensionality reduction and Tabu Search among others.

24.6.1 Direct Methods

Various algorithms have been applied in literature to solve mathematical programming formulations directly for Integer Programming (IP) and Dynamic Programming (DP) formulations. In [34] authors claim that mathematical programming optimization methods such as IP and DP suffer from the curse of dimensionality and they cannot deal with nonlinearity and non-differentiability of objective functions. Authors using DP focus on reducing the computation time required. Authors using IP use various branch and bound type methods to solve their formulations.

24.6.1.1 Integer Programming

IP formulations are generally solved using enumeration techniques such as branch and bound algorithms. In [11] LINGO is used to model and solve the mixed integer programming problem. Authors in [29, 30, 33], also solve directly using GAMS software and [42] solves directly using PP/MS program. The author in [32] uses the simplex method to solve the 0/1 IP problem.

In [38] the authors used an extension of an implicit enumeration technique using a decision tree for the whole system where each branch is a solution; this method is combined with a heuristic method. In [35] a decision tree based approach is also used. A branch and bound algorithm is also applied in [51].

24.6.1.2 Dynamic Programming

A number of papers acknowledge that the computational time required for DP optimization is a disadvantage for this approach. In [47] the author agrees that the limitation for DP is the computational time.

Approaches have been proposed to deal with the problem of large computation times for this approach. In [47] a dimensionality reduction method, DPSA (Dynamic Programming by Successive Approximations), is applied to deal with the large computation time. The algorithm then iteratively keeps a set of solutions constant and then optimizes the remaining variables.

In [21] the author notes that the calculations do not need to occur in real time. The fuzzy dynamic programming approach is applied with objective functions and constraints replaced by membership values.

24.6.1.3 Heuristic Methods

Heuristic methods have been applied in a number of studies, as they use a rule of thumb type approach to search for a solution [42]. In [17] the author suggests that heuristic methods can solve some of the limitations of mathematical programming methods by splitting the problem into units and then scheduling activities sequentially. In [9] authors note that heuristics require significant operator input and may not even find solutions which are feasible. Although [14] states that literature indicates that heuristics are still widely in use.

In [38] a heuristic method called fathoming is used to search through the branches until each branch can no longer contain optimal solutions. In [41] a heuristic maintenance class based approach is applied to perform corrective and preventive maintenance to achieve cost effective maintenance intervals.

24.6.2 Artificial Intelligence Methods

A variety of Artificial Intelligence (AI) approaches have been applied in literature as alternatives to heuristic methods. For example Genetic Algorithms (GAs) are applied by 7 studies.

GAs are a well-established population based evolutionary search method. Various studies attempt to reduce the computation time associated with GAs. Authors in [17] propose modifications to the GA such as string reversal and reciprocal exchange mutation. The modifications proposed are shown to improve the computational performance of the algorithm and perform better than DP. In [18] the authors attempts to reduce the computation time by proposing a code specific constraint transparent process to deal with the heavy computation time. The results show improved computation time compared to binary GA approaches. In [16] the author uses the NSGAII algorithm to perform multi-objective optimization. This approach ensures reliability and power economy are both maximized simultaneously. In [39] the author proposes an Advanced Progressive Real Coded Genetic Algorithm (APRCGA), where the chromosomes consist of real numbers which are later converted into integers to represent solutions. In [37] authors apply a GA approach to move outages from periods of low reliability to high reliability.

Other evolutionary computation methods have been applied such as Simulated Annealing (SA) and Particle Swarm Optimization (PSO); the results have been compared to empirical and exact results to assess performance. In [25] a multi stage SA optimization process is applied; the approach was compared to an enumeration method with a near global optimal solution being found. Authors in [36] also apply

SA to compare different cooling schedules, neighbourhood move operators and hybrid approaches in an attempt to improve the traditional SA. Authors in [23] propose a non-periodic PSO based approach and use continuous values within the search space. In [52] a modified PSO (MPSO) algorithm is used to reduce the number of control parameters for the algorithm. Authors in [53] apply a revised PSO with capability to deal with inequality and equality constraints using a penalty function. An Improved Binary Particle Swarm Optimization (IBPSO) method is applied in [24] to avoid premature convergence and to improve the search quality of the traditional PSO.

Other methods include Ant Colony Optimization (ACO) algorithm [48], a clonal selection algorithm based on the Artificial Immune System [46] and a Teaching Learning Based Optimization algorithm [45].

24.6.3 Hybrid Methods

Hybrid methods have been increasingly applied in recent years; these methods combine the advantages of different approaches such as AI and heuristic methods. The proposed hybrid methods are generally found to outperform existing methods.

Hybrid methods involving SA are particularly popular. Authors in [20] compare GA and GA/SA hybrid approaches; the hybrid method produces the fastest convergence speed and both methods are shown to lead to the same results. In [19] authors also use a GA/SA hybrid approach using an encoding and decoding method. This method is shown to create remarkably shorter computation times than the SA approach and produce better convergence and results than the GA approach. In [54] it is found that by applying the GA/SA hybrid approach each individual solution can be improved by SA, and similarly to [19] the convergence of the GA/SA algorithm is superior to the convergence of the GA method alone. Investigation of hybrid local search and SA method in [36] finds that the solution quality can be improved through application of this hybrid method.

In [34] the authors apply a Hill Climbing Technique (HCT) and a hybrid Extremal Optimization and GA method (EO/GA) and compare the results. The EO/GA technique is shown to be superior in both cases with respect to the average objective function values. In [43] results are compared from a hybrid PSO/GA method and a PSO/Shuffled Leap Frog algorithm with other methods; the conclusion is that the Shuffled Leap frog hybrid is an efficient and robust optimization strategy.

A different combination of DP formulation and GAs is applied in [21]. This combination of dynamic programming, GAs and fuzzification is shown to produce optimal maintenance schedules for a given test case. In [15] an evolutionary programming method is applied to find a near optimal solution and subsequently a Hill-Climbing method is used to ensure the feasibility of solutions is maintained.

24.6.4 Other Methods

Several other optimization techniques have been identified and used within studies to solve scheduling optimization problems.

Deterministic solvers have been considered in the place of stochastic AI methods such as GAs. In [26] the authors use a Tabu Search algorithm for scheduling of outage tasks. This approach has a fast search speed and produces diversified solutions avoiding local optima.

Approaches which consider the problem to be continuous instead of discrete have also been considered. In [55] a sequential continuous approach to the gas turbine maintenance problem is proposed. The results indicate decreasing maintenance intervals as the turbine ages to ensure reliability.

Other approaches have also been considered within studies. Lagrangian Relaxation is applied in [49] to decompose the problem into tractable scenarios. In [4] a dimensionality reduction method is used to compare maintenance intervals and replacement rates.

24.7 Discussion

Through analysis of the chosen papers, a number of key deductions and trends can be identified; these are detailed as follows in section order.

The first trend identified was a significant increase in interest over the recent years. Figure 24.3 shows an increase in the number of papers produced per decade, indicating a strong increase in research in optimization of power plant preventive maintenance scheduling over recent years. This review focusses on papers where different optimization methods are applied; hence it can be seen that there is more research in application of different optimization methods in recent years.

Analyzing problem formulations has led to some key observations. The most common objective function considered is the cost objective function; this includes optimization of profit and maintenance and operational costs. Studies also combine and compare these objectives [5, 27]. Reliability is considered to be a vital objective function and the levelling of the reserve rate is used by a number of papers to ensure that demand is met at all times [9, 33, 34]. The conflict between reliability and cost has been acknowledged; it is more significant when the maintenance and operation are carried out by different companies [52]. The conflict between these objectives has been explicitly dealt with in a number of papers, which combine and compare results for these objective functions.

The most common constraint used within literature is the manpower constraint as this involves costs; it is often formulated as a soft constraint enabling solution flexibility. On the other hand technological constraints such as durations are hard constraints as these are formulated based on manufacturer requirements. Constraints are also used by some papers to compare 2 main objective functions, where one is set as an objective and the other as a constraint and vice versa. The application of constraints to the optimization problem is shown to reduce the objective function values in a number of papers [30, 42, 49].

The vast majority of papers formulate the model as mathematical programming models. IP and DP in particular are the main formulations for the MS problem.

An increase in the use of stochastic methods to evaluate solutions, as opposed to deterministic methods, has also been observed. In particular, in [50] a growing importance given to the explicit treatment of the non-deterministic parameters is noted. In [20] deterministic and stochastic methods are compared and it is concluded that stochastic methods can evaluate the risk involved with deteriorating equipment. A few other studies have begun to use stochastic parameters in some form [4, 5, 35].

The increasing use of multi-objective optimization and hybrid optimization methods has been observed. In [13] it is identified that the literature of multi-objective scheduling is much sparser than that of single objective scheduling and that since 1995 there has been an increasing interest in the area. For instance, in [50] a multi-objective optimization approach with the branch and bound method is applied, in [17] authors use a GA with two objectives. Hybrid methods have also been increasingly applied in recent years [19, 20, 54]. These areas of multi objective and hybrid optimization are an interesting area for future research.

A research gap has been identified in the use of stochastic methods to model the MS problem. Discrete Event Simulation has not been used for modelling the MS within the power industry although it has been applied within other industries. The explicit treatment of uncertain parameters using Discrete Event Simulation could be an interesting application. The application of Discrete Event Simulation with multi objective hybrid optimization is a novel and exciting area for research. These prospective areas for research could provide results with a number of benefits for power generation companies.

24.8 Conclusions

This study has reviewed papers in preventive maintenance scheduling of equipment for power plants. This paper presents up to date detailed review of papers within power plant preventive maintenance scheduling with details of objective functions, constraints, model formulation and optimization methods. Significant conclusions, trends and a research gap have been identified from the literature. Trends such as the increasing use of stochastic parameters and a growing use of hybrid methods and multi objective optimization have been noted. Research has not been carried out to apply Discrete Event Simulation with optimization to power plant preventive maintenance scheduling. Combining these various research gaps could generate interesting results and revelations. These are excellent opportunities to be exploited in future research.

References

1. Xu B, Han XS, Li M, Xiao DL (2012) System maintenance scheduling: review and prospect. 2012 IEEE innovative smart grid technologies—Asia, ISGT Asia 2012
2. Camci F (2009) Comparison of genetic and binary particle swarm optimization algorithms on system maintenance scheduling using prognostics information. Eng Optim 41(2):119–136
3. Yamayee ZA (1982) Maintenance scheduling-description, literature survey, and interface with overall operations scheduling. IEEE Trans Power Appar Syst V PAS-101(N 8):2770–2779
4. Katafuchi T, Nakamura M, Suzuki Y, Hatazaki H (2003) Proper decision for maintenance intervals of equipment in thermal power station based on operational rate and maintenance replacement rate. Electr Eng Jpn (English translation of Denki Gakkai Ronbunshi) 145(1):10–19
5. Zhao Y, Volovoi V, Waters M, Mavris D (2006) A sequential approach for gas turbine power plant preventative maintenance scheduling. J Eng Gas Turbines Power 128(4):796–805
6. Greasley A (2000) Using simulation to assess the service reliability of a train maintenance depot. Qual Reliab Eng Int 16(3):221–228
7. Oyarbide-Zubillaga A, Goti A, Sanchez A (2008) Preventive maintenance optimisation of multi-equipment manufacturing systems by combining discrete event simulation and multi-objective evolutionary algorithms. Prod Plan Control 19(4):342–355
8. Anne-Sylvie C, Floru I, Azzaro-Pantel C, Pibouleau L, Domenech S (2003) Optimization of preventive maintenance strategies in a multipurpose batch plant: application to semiconductor manufacturing. Comput Chem Eng 27(4):449–467
9. Dahal KP, Aldridge CJ, McDonald JR (1999) Generator maintenance scheduling using a genetic algorithm with a fuzzy evaluation function. Fuzzy Sets Syst 102(1):21–29
10. Liu Y, Meng X, Sheng W (2011) Research and application of maintenance schedule optimization based on intelligent algorithm. APAP 2011—proceedings: 2011 international conference on advanced power system automation and protection, vol 2, p 957
11. Alardhi M, Hannam RG, Labib AW (2007) Preventive maintenance scheduling for multi-cogeneration plants with production constraints. J Qual Maint Eng 13(3):276–292
12. Maccarthy BL, Liu J (1993) Addressing the gap in scheduling research: a review of optimization and heuristic methods in production scheduling. Int J Prod Res 31(1):59–79
13. Lei D (2009) Multi-objective production scheduling: a survey. Int J Adv Manuf Technol 43 (9–10):925–938
14. Kralj BL, Petrovic R (1988) Optimal preventive maintenance scheduling of thermal generating units in power systems: a survey of problem formulations and solution methods. Eur J Oper Res 35(1):1–15
15. El-Sharkh MY, El-Keib AA (2003) Maintenance scheduling of generation and transmission systems using fuzzy evolutionary programming. IEEE Trans Power Syst 18(2):862–866
16. Zhang D, Li W, Xiong X (2012) Bidding based generator maintenance scheduling with triple-objective optimization. Electr Power Syst Res 93:127–134
17. Baskar S, Subbaraj P, Rao MVC, Tamilselvi S (2003) Genetic algorithms solution to generator maintenance scheduling with modified genetic operators. IEE Proc: Gener Trans Distrib 150(1):56–60
18. Wang Y, Handschin E (2000) New genetic algorithm for preventive unit maintenance scheduling of power systems. Int J Electr Power Energy Syst 22(5):343–348
19. Kim H, Nara K, Gen M (1994) A method for maintenance scheduling using GA combined with SA. Comput Ind Eng 27(1–4):477–480
20. Mohanta DK, Sadhu PK, Chakrabarti R (2007) Deterministic and stochastic approach for safety and reliability optimization of captive power plant maintenance scheduling using GA/SA-based hybrid techniques: a comparison of results. Reliab Eng Syst Saf 92(2):187–199
21. Huang S (1998) A genetic-evolved fuzzy system for maintenance scheduling of generating units. Int J Electr Power Energy Syst 20(3):191–195

22. Suresh K, Kumarappan N (2013) Coordination mechanism of maintenance scheduling using modified PSO in a restructured power market. In: Proceedings of the 2013 IEEE symposium on computational intelligence in scheduling, CISched 2013—2013 IEEE symposium series on computational intelligence, SSCI 2013, p 36
23. Pereira CMNA, Lapa CMF, Mol ACA, Da Luz AF (2010) A particle swarm optimization (PSO) approach for non-periodic preventive maintenance scheduling programming. Prog Nucl Energy 52(8):710–714
24. Suresh K, Kumarappan N (2013) Generation maintenance scheduling using improved binary particle swarm optimisation considering aging failures. IET Gener Transm Distrib 7 (10):1072–1086
25. Yokoyama R, Niimura T (1996) Thermal generating unit maintenance scheduling by multi-stage application of simulated annealing. Proc IEEE Int Conf Syst Man Cybern 2:1531
26. Sawa T, Furukawa T, Nomoto M, Nagasawa T, Sasaki T, Deno K, Maekawa T (1999) Automatic scheduling method using tabu search for maintenance outage tasks of transmission and substation system with network constraints. IEEE engineering society, winter meeting, vol 2, p 895
27. Cao J, Bell KRW, Kockar I, San Martin LAS (2010) Generation maintenance scheduling in a liberalised electricity market. Proceedings of the universities power engineering conference
28. Perez-Canto S (2011) A decomposition approach to the preventive maintenance scheduling problem: an empirical example for the Spanish power plant system. Int J Manuf Technol Manage 22(1):58–77
29. Perez-Canto S, Rubio-Romero JC (2013) A model for the preventive maintenance scheduling of power plants including wind farms. Reliab Eng Syst Saf 119:67–75
30. Badri A, Niazi AN, Hoseini SM (2012) Long term preventive generation maintenance scheduling with network constraints. Energy Procedia 14:1889
31. Marwali MKC, Shahidehpour SM (1998) A deterministic approach to generation and transmission maintenance scheduling with network constraints. Electr Power Syst Res 47 (2):101–113
32. Perez-Canto S (2011) Using 0/1 mixed integer linear programming to solve a reliability-centered problem of power plant preventive maintenance scheduling. Optim Eng 12(3):333–347
33. Perez-Canto S (2008) Application of Benders' decomposition to power plant preventive maintenance scheduling. Eur J Oper Res 184(2):759–777
34. Reihani E, Najjar M, Davodi M, Norouzizadeh R (2010) Reliability based generator maintenance scheduling using hybrid evolutionary approach. 2010 IEEE international energy conference and exhibition, EnergyCon 2010, p 847
35. Abiri-Jahromi A, Fotuhi-Firuzabad M, Parvania M (2012) Optimized midterm preventive maintenance outage scheduling of thermal generating units. IEEE Trans Power Syst 27 (3):1354–1365
36. Schlünz EB, Van Vuuren JH (2013) An investigation into the effectiveness of simulated annealing as a solution approach for the generator maintenance scheduling problem. Int J Electr Power Energy Syst 53(1):166–174
37. Eshraghnia R, Mashhadi HR, Shanechi MHM, Karsaz A (2005) A novel approach for maintenance scheduling of generating units in a competitive environment. 2006 IEEE power India conference, vol 2005, p 821
38. Dopazo JF, Merrill HM (1975) Optimal generator maintenance scheduling using integer programming. IEEE Trans Power Appar Syst PAS-94(5):1537–1545
39. Aghaie M, Norouzi A, Zolfaghari A, Minuchehr A, Mohamadi Fard Z, Tumari R (2013) Advanced progressive real coded genetic algorithm for nuclear system availability optimization through preventive maintenance scheduling. Ann Nucl Energy 60:64–72
40. Mohammad Hadi Hadavi S (2009) A heuristic model for risk and cost impacts of plant outage maintenance schedule. Ann Nucl Energy 36(7):974–987
41. Adhikary DD, Bose GK, Bose D, Mitra S (2013) Maintenance class-based cost-effective preventive maintenance scheduling of coal-fired power plants. Int J Reliab Saf 7(4):358–371

42. Mukerji R, Merrill HM, Erickson BW, Parker JH, Friedman RE (1991) Power plant maintenance scheduling: optimizing economics and reliability. IEEE Trans Power Syst 6 (2):476–483
43. Giftson Samuel G, Christober Asir Rajan C (2015) Hybrid: particle swarm optimization-genetic algorithm and particle swarm optimization-shuffled frog leaping algorithm for long-term generator maintenance scheduling. Int J Electr Power Energy Syst 65:432–442
44. Fattahi M, Mahootchi M, Mosadegh H, Fallahi F (2014) A new approach for maintenance scheduling of generating units in electrical power systems based on their operational hours. Comput Oper Res 50:61–79
45. Abirami M, Ganesan S, Subramanian S, Anandhakumar R (2014) Source and transmission line maintenance outage scheduling in a power system using teaching learning based optimization algorithm. Appl Soft Comput J 21:72–83
46. El-Sharkh MY (2014) Clonal selection algorithm for power generators maintenance scheduling. Int J Electr Power Energy Syst 57:73–78
47. Yamayee ZA, Sidenblad K (1983) Computationally efficient optimal maintenance scheduling method. IEEE Trans Power Appar Syst PAS-102(2):330–338
48. Vlachos A (2013) Ant colony system algorithm solving a thermal generator maintenance scheduling problem. J Intell Fuzzy Syst 24(4):713–723
49. Wu L, Shahidehpour M, Li T (2008) GENCO's risk-based maintenance outage scheduling. IEEE Trans Power Syst 23(1):127–136
50. Kralj B, Rajaković N (1994) Multiobjective programming in power system optimization: new approach to generator maintenance scheduling. Int J Electr Power Energy Syst 16(4):211–220
51. Kralj B, Petrovic R (1995) A multiobjective optimization approach to thermal generating units maintenance scheduling. Eur J Oper Res 84(2):481–493
52. Kumarappan N, Suresh K (2005) Coordinated maintenance scheduling in power system using combined genetic algorithm and simulated annealing. 2006 IEEE power India conference, vol 2005, p 416
53. Ekpenyong UE, Zhang J, Xia X (2012) An improved robust model for generator maintenance scheduling. Electr Power Syst Res 92:29–36
54. Leou R (2006) A new method for unit maintenance scheduling considering reliability and operation expense. Int J Electr Power Energy Syst 28(7):471–481
55. Zhao Y, Volovoi V, Waters M, Mavris D (2005) A sequential approach for gas turbine power plant preventive maintenance scheduling. Proceedings of the ASME power conference, 2005, vol PART A, p 353

Chapter 25
Beyond RAMS Design: Towards an Integral Asset and Process Approach

A. Martinetti, A.J.J. Braaksma and L.A.M. van Dongen

Abstract The lifespan to which assets can be efficiently maintained, upgraded or disposed, heavily depends on the characteristics designed into the asset in the design phase. RAMS analysis is a well-established approach often used to reach this target. This approach is however not adequate for handling the complexity of changes and demands placed on nowadays assets. This can lead to reduced performance and unnecessary risk taking. There is a need for a more integral RAMS (SHEEP) perspective including Supportability, Health, Environment, Economics and Politics. Additionally there is often only focused on the asset itself and not on processes supporting the maintenance of an asset. Therefore this chapter does not only give a historic overview on RAMS evolvement, but also aims at answering how the supporting processes can be designed from an integral RAMSSHEEP perspective. We illustrate this by analysing the functional requirements for the Toilet System (TS) of the Sprinter Light Train (SLT).

25.1 Introduction

The product's Reliability, Availability, Maintainability and Safety (RAMS) are very important characteristics that have to be embedded and designed into every product or asset. This is because design decisions have a large influence on these characteristics. For example the expenditures are made during the equipment's life time, but those costs are already, for a large part, committed during the development stage of the equipment's life cycle. A re-design of components and parts during the later life cycle stages can have a strong influence on the maintenance efforts [18].

A. Martinetti (✉) · A.J.J. Braaksma · L.A.M. van Dongen
Universiteit Twente. Drienerlolaan 5, 7522 NB Enschede, The Netherlands
e-mail: a.martineeti@utwente.nl

L.A.M. van Dongen
e-mail: l.a.m.vandongen@utwente.nl

L.A.M. van Dongen
Netherlands Railways, Stationshal 15, 3511 CE Utrecht, The Netherlands

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_25

Previous research already pointed out the importance of considering maintenance in planning the product life cycle [24], RAMS is considered a strategical, tactical and operational, risk-driven maintenance concept, in which a system's or asset's Reliability, Availability, Maintainability and Safety have to be taken into account. RAMS is applied to facilitate competitive advantage for the product and to reduce the business risk associated with non-performance of products and systems [16]. The approach helps designer and analysts to get an indication of the performance of the functioning of a system.

As described also by the European Committee for Electrotechnical Standardization (CENELEC) [9], RAMS is "a qualitative and quantitative indicator of the degree that the system, or the subsystems and components comprising that system, can be relied upon to function as specified and to be both available and safe".

But nowadays, this vision is not more enough and additional step is required to address environmental and political requests at the same time. BCG [17] states that "business is on the verge of a major 'next wave' of asset productivity improvement —one that will go farther and be more difficult to achieve than past initiatives", identifying such as the exhaustion of traditional cost cutting. This challenge can be found in complex systems such as a railway system where transportation performance cannot be guaranteed just by technically perfect design concepts, but where the results are heavily affected by specific procedures, working regulations and working conditions. [21].

In the following sections we will first describe the elements of a full RAMSSHEEP methodology, then we take the method in a historic perspective, thirdly we will illustrate the application of RAMSSHEEP at our case company (NS) and finally we will present our conclusions.

25.2 RAMSSHEEP Methodology

The decision process represent a not negligible problem in terms of time and money when the product is a capital asset since the required services to provide support have to be determined for its entire life cycle [19]. Therefore, the evaluation decisions should include making decisions about not only the asset but also about the ancillary activities that it requires.

As mentioned earlier the RAMSSHEEP methodology aims to connect the well-known aspects of the RAMS approach with five essential parameters (Supportability, Health, Environment, Economics and Politics) in order to design, plan, realise, use and dispose an asset increasing the efficiency and reducing costs and environmental impacts [4]. Table 25.1 organizes and defines the nine elements of the methodology, pinpointing with the help of some sources the main characteristics of each point.

The nine described elements of the RAMSSHEEP approach can be divided in three macro-categories as represented in Fig. 25.1.: (1) RAMS (Reliability, Availability, Maintainability, Supportability) concerning the aspects related to

Table 25.1 Elements of the RAMSSHEEP methodology

Element	Definition	Contextualization
Reliability	"The probability that an asset can perform a required function under given conditions for a given time interval" [23]	The reliability of a train is for example 90%. This means that there is a certainty of 90% that the train could travel
Availability	"The ability of an asset to be in a state to perform a required function under given conditions at a given instant of time assuming that the required external resources are provided" [20]	The availability of a train is for example 85%. This means that the train should be operational circa 310 days/year
Maintainability	"The probability that following the occurrence of a failure of an asset will once again be operational within a specific time" [23].	The maintainability of a train is for example 90%. This means that there is a certainty of 90% that the train will be put in service on time after a maintenance action[1]
Supportability	"The characteristic of an asset to influence the easiness with which logistic resources can be available at the right time at the right place" [18]	The supportability of an asset can heavily affect the logistic organization causing delays (waiting for spare parts, technicians, equipment available) during the maintenance operations and influencing the Mean Time To Maintain (MTTM)
Safety	"A state in which or a place where you are safe and not in danger or at risk" [3]; "Freedom from unacceptable risks of harm" [9]	The Safety has to be included to ensure a safe asset for the final users and safe working places for the personnel involved in the production and in the maintenance operations. To note, how the absence of safety could change the cost-effectiveness of an asset
Health	"Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" [27]	Health has to be included to ensure that an asset does not cause diseases for the final users and for the personnel involved in the production and in the maintenance operations
Environment	"The environment represents the earth, including rocks, soils, water, air, atmosphere and living things" [12]	The asset should reduce as much as possible, for example by using the Best Available Techniques (BAT) [10] the impact on the Environment during the entire life-cycle. Here lies the difference between environmental compatibility and environmental sustainability

(continued)

Table 25.1 (continued)

Element	Definition	Contextualization
Economics	"The economic perspective is concerned with the financial aspects of the asset and its operation." [22]	The economic factors often drive the main direction and the investment from the design phase to the decommission phase of a product/asset
Politics	"The first definition of politics was used in the Aristotle's book Πολιτικά, Politika, referring to the *affairs of the cities*"	The politic decisions should affect the main direction of a capital assets investment pinpointing and underlining the needs of the community

[1] To note that, in addition to the stochastic definition, the Maintainability could also represent the level of easiness to maintain an asset/product/component. In other words, how quickly maintenance activities can be performed reaching the required level of quality

maintenance management, reliability, logistic and spare parts, (2) SHE regarding all the criticalities that could cause injuries, fatalities, diseases and environmental disasters during the design, production operations, management and decommission and (3) EP including all the political and economic considerations and evaluations on the feasibility and need of the project for the society and, in general, for the market.

Adopting an automotive metaphor, the macro-categories have to work as the elements in a gearbox, providing the car requests (high torque when climbing hills and when starting at low speeds and low torque running at high speeds on level roads due the inertial momentum) at the right moment according to the situation's conditions. In the same way, RAMS, SHE, EP gears should provide during the design phase precise and essential information over the impacts of the project on the different aspects ensuring to reach as much as possible a design-effectiveness of the production system.

25.3 R, RM, RAMS, RAMS-LCC and RAMSSHEEP: Placed in Historic Perspective[1]

Looking back to the last six decades, we can identify a clear evolution of the dedicated approaches to ensure that products/assets could perform a required function under given conditions for a given time interval. The Fig. 25.2 should help to resume the observed improvements and changes, pinpointing how the demands moved from an exclusive product perspective (R, RM), to the awareness to ensure

---

[1] R (Reliability), RM (Reliability, Maintainability), RAMS (Reliability, Availability, Maintainability and Safety), RAMS-LCC (Reliability, Availability, Maintainability Safety—Life Cycle Cost).

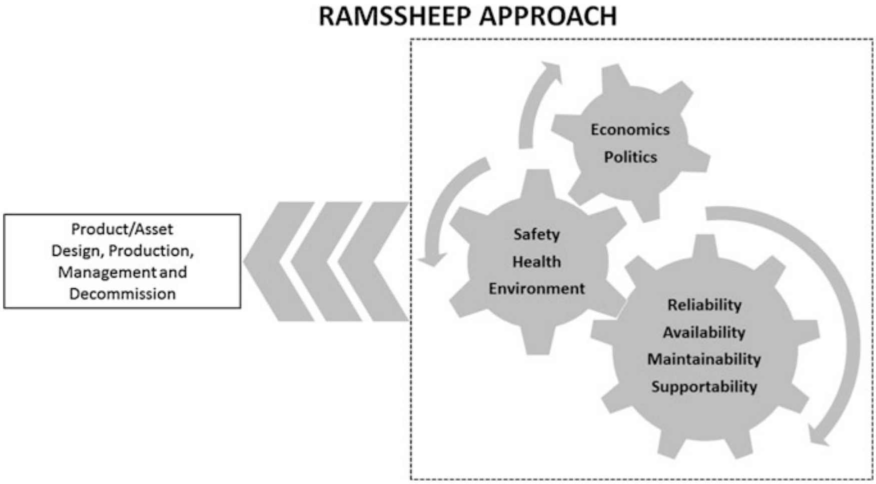


Fig. 25.1 RAMSSHEEP: RAMS, SHE and EP elements working together as parts of a gearbox

safe products for workers and users (RAMS) to cost-optimization (RAMS-LCC) products, towards the need to ensure society approval and value (RAMSSHEEP).

The first structured approach of the Modern Age to evaluate the reliability in the industrial production process was introduced after World War Two in 1954 [13] during the First National Symposium on Reliability and Quality Control in the United States of America (US). The need of reliability was mainly fuelled by two different but connected events.

During the Second World War (i), 60% of the airborne equipment and spare parts arrived damaged and unserviceable before use, causing a remarkable waste of resources, energy, money, time and man-working hours. Due to (ii) the technological evolution, the complexity of the system and the number of system components increased (and it is still increasing), directly affecting the reliability of the entire system as shown in the example of Table 25.2 with a simple example. To better describe the importance of this phenomena, even more actual nowadays, the authors want to underline the different approach in terms of number of components used to design and build the Boeing 747 and the Mariner/Mars ‘64. The aircraft was assembled using more than 4.500.000 parts [1]; but for the spacecraft, since the success of the mission was strictly dependent on the reliability of every part, only 138.000 components [8] were used during the construction in order to reduce the number of unnecessary and unreliable elements.

In the ‘80s due to the high products demand, the necessity to increase the maintainability performance reducing the downtimes related to parts replacement and repairing offered the opportunity to dedicate more efforts on the design for maintenance aspects. The manufactures invested in using materials that did not prolong maintenance activities, using standard and universal applicable components, fasteners to accelerate maintenance activities, providing sufficient space

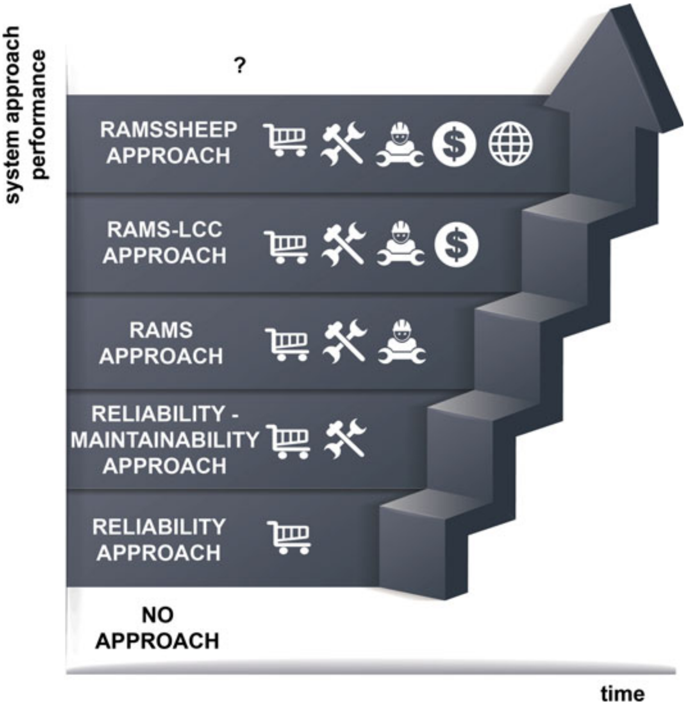


Fig. 25.2 Evolution of the approaches for product/asset production

Table 25.2 Reliability and increasing product complexity [14]

Farm tractor model year	Number of critical components	Tractor reliability per year[1] [%]	Number of tractors failing per year/1000 tractors
1935	1.200	88.7	113
1960	2.250	79.9	201
1970	2.400	78.7	213
1980	2.600	77.1	229
1990	2.900	74.8	251

[1] Assuming an average component reliability of 99.99% and critical components reliability-wise in series

around the maintenance points and designing equipment in such a way that it can only be maintained in the right way. Following the innovative vision called "Prevention through Design (PtD)" in terms of Occupational Health and Safety (OHS) proposed by the National Institute for Occupational Safety and Health (NIOSH) in the US, at the end of '80s the European Community decided to introduce several directives [5, 6], in order to oblige the employers to carry out an exhaustive Risk Assessment and Management analysis to provide safe and reliable machineries equipment, products and workplaces. Moreover, as a general approach, in the European Directives, the maintenance workers were identified as "workers who may be at increased risk", so that the need to conduct a separate Risk Assessment and Management for the maintenance activities becomes more evident and necessary [15]. The aforementioned evolution helps to embrace a more general approach already explained and well-known as RAMS. Not only the products/assets and their maintenance characteristics, but also the occupational conditions of the workers have to be taken into account.

Meanwhile, in an attempt to improve the design of products and reduce design changes, cost, and time to market, concurrent engineering or Life Cycle Engineering (LCE) was emerging as an effective approach to addressing these issues in competitive global market [2]. Dowlatshahi [7] underlined that the design of the product influences between 70% and 85% of the total cost of a product remarking how the designers have the opportunity to substantially reduce the Life Cycle Cost (LCC) of the product. As happened for the Reliability approach, the first motivation and incentive were provided by the weapons' market, stimulated by Department of Defence in order to reduce the operations and support cost that were accounted for 75% of the total expense [11].

Lastly, the development of the "Green Economy" theory defined as "low carbon, resource efficient, and socially inclusive (where) growth in income and employment should be driven by public and private investments that reduce pollution, enhance energy and resource efficiency, and prevent the loss of biodiversity and ecosystem services" [25], forced to include environmental awareness to the production system to decrease carbon emissions and ecological footprint.

25.4 Design for RAMSSHEEP: A Case Study on an Existing Capital Asset (Toilet System) in Dutch Railways

To illustrate the opportunities offered by the RAMSSHEEP methodology on a real case, a test study on a capital asset is provided. The Dutch Railways (Nederlandse Spoorwegen, NS) have to place new Toilet Systems (TS) in the Sprinter Light Trains (SLT) not designed to have them due to the very short distance to run between two stations for which they were projected. Unfortunately, after a first working period the public opinion complaints forced the owner to rethink the first decision starting an evaluation process in order to add the TS.

The goal of the application is to devise the implementation of TS into the trains using RAMSSHEEP-principles which should lead to design criteria helping to focus on alternative design solutions.

The importance of a careful, detailed and integral life-cycle design of the TS for the SLT is also enhanced by the failure data provided by the asset manager. As stated in [26], the 13% of the failures of the rolling stock are attributed by technical problems occurred to the TS.

To provide and create a brief impression about the possible layout of the major TS sub-systems the analyses are also coupled to some design requirements (considered as pre-requirements by the asset owner and not included in the RAMSSHEEP approach):

- TS should be situated at the multifunctional vestibule;
- TS should be equipped with a vacuum toilet system which will be connected to a biological wastewater treatment system (bioreactor);
- The semi-closed wastewater treatment system should process the human waste in order to separate solids and fluids that will be biologically treated with aerobic and anaerobic processes.

25.4.1 RAMSSHEEP Principles Applied on the TS

Since NS requires a high-standard system, a long detailed list of technical specifications is analysed in order to match them with the principles of the RAMSSHEEP methodology.

The full list of the most important requirements is composed by 127 functional elements which were arbitrarily categorized by the students involved in the process. This analysis showed that besides the RAMS (57%) principles, the SHEEP (43%) principles represents an important part to be included in the design of an asset (R (20%), A (5%), M (10%), S (22%), S (16%), H (9%), E (7%), E (2%) and P (9%)).

An example of technical specifications of the SLT TS, as provided by NS, are shown according to each specific principle (Table 25.3). This approach gives the opportunity to consider in the evaluation every aspect of the problem in a “future-proof” vision.

To create a connection between the RAMSSHEEP approach and pre-requirements, the TS is divided into five subsystems (outside design, duty system, water system, sensor system, and personal care) during the design phase in order to make the design and the design choices as more efficient as possible.

Moreover, during the design process of the subsystems the users, cleaners and maintenance technicians were taken into account in order to create a TS able to be maintained and accessible from the inside regardless of the location, such as a train station or a workshop.

Table 25.3 An example of RAMSSHEEP principles applied to a few of the supplied TS requirements

Reliability	Availability	Maintainability
The toilet should not have negative influence on the reliability of existing systems and should guarantee a MTBF of at least 40.000 h. The lifespan needs to be at least 30 years	The toilet, including the waste water treatment system and the fresh water system, should have a capacity sufficient for the number of passengers	Components of the toilet should be positioned in such a way that it provides easy access for the maintenance procedures
Safety	Supportability	Health
The toilet should offer an adequate level of safety in the event of a collision or derailment during normal service	The toilet should use standard, universally applicable components, tools and parts	The toilet should be easy to clean since people would be reluctant to use them if they are dirty
Environment	Economics	Politics
Application of hazardous waste on the product should be avoided as well as contaminating chemicals which can impact the environment	The LCC of the toilet system should be within the financial boundaries	The toilet should be installed even in trains meant for short commuting

The combination of technical specifications and pre-requirements with the RAMSSHEEP approach gives the advantage to design a concept solution able to fulfil the most important aspects in order to have a Reliable, Available, Maintainable, Safe, Supportable, Healthy, Environment-friendly and Economic and Politic-feasible system as shown in Fig. 25.3.

Therefore, the minimal and simple design created with the RAMSSHEEP principles offers several advantages from a long term prospective. Most of the maintenance tasks are carried out on a human-handling level and do not involve lifting heavy parts (waste bags could be dragged over the floor, the dispensers and hand dryers could be cleaned avoiding un-ergonomic positions). The technicians should be able to solve most of the failures without docking the train at the workshop. The outside hatches should permit accessing the bio filter, chemical box and water reservoirs during extensive maintenance periods. The technicians could access the electricity and water reservoir panels from the inside in order to fix problems quickly during unexpected downtime situations.

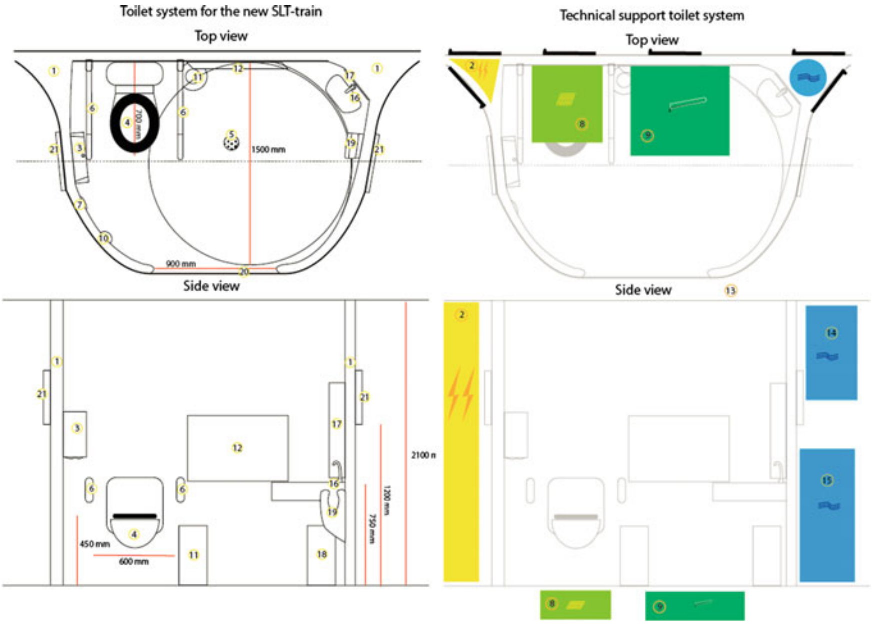


Fig. 25.3 Overview of complete design concept (1. Outer casing, 2. Electricity panel, 3. Toilet paper, 4. Toilet bowl, 5. Drain, 6. Armrests, 7. Flush button, 8. Bio-filter, 9. Chemical box, 10. Help-button, 11. Hygiene box, 12. Nursery, 13. Heating and air-conditioning, 14. Water reservoir tap, 15. Water reservoir toilet, 16. Tap, soap dispenser, 17. Mirror, 18. Bin, 19. Hand dryer, 20. Door, 21. Information screens)

25.5 Conclusions and Recommendations

Our literature review showed that the RAMS(SHEEP) Methodology has evolved from purely technically centered (Reliability and Availability) to incorporate also Maintainability and Safety concepts. Due to quicker changing environments and more critical demands (e.g. legislation and political influences) there is a need for taking additional factors into consideration. The exemplary case of the Toilet System at NS clearly illustrates that requirements nowadays should include SHEEP factors besides the well-known RAMS. The weighting of the individual elements in the design of assets is moving from a pure technical and cost perspective to a more value based evaluation. RAMSSHEEP analysis is primarily aimed at the design phase of an asset but has potential to be used during the entire life-cycle management. Further research should be performed on how to re-use and update existing analyses during the life-cycle.

Acknowledgements this research is partly funded by the Lloyd's Register Foundation (LRF) supporting engineering-related education, public engagement and the application of research. The authors thank NS for their cooperation. Authors also acknowledge the master students of the course in "Design for Maintenance Operations" a.y. 2015–2016 at the University of Twente, for inspiring them with interesting suggestions and analysis provided during the lectures.

References

1. Apple FC (1970) The 747 Ushers in a new era, the American way, American airlines magazine, March 1970, pp 25–29
2. Asiedu Y, Gu P (1998) Product life cycle cost analysis: state of the art review. Int J Prod Res 36(4):883–908. doi:10.1080/002075498193444
3. Cambridge International Dictionary of English (2013) ISBN-13: 978-0521484213
4. Charles A, Schuman A, Brent C (2005) Asset life cycle management: towards improving physical asset performance in the process industry. Int J Oper & Prod Manag 25(6):566–579. doi:10.1108/01443570510599728
5. Council of the European Communities (1989) Council directive 89/391 EEC concerning the introduction of measures to encourage improvements in the safety and health of workers at work
6. Council of the European Communities (1989) Council directive 89/392/EEC on the approximation of the laws of the member states relating to machinery
7. Dowlatshahi S (1992) Product design in a concurrent engineering environment: an optimization approach. J Prod Res 30(8):1803–1818
8. Engineering Opportunities (1964) Every aspect of Mariner/Mars '64 missions involves basic state-of-art advances, October 1964, pp 14–15
9. European Committee for Electrotechnical Standardization-CENELEC (1999) Railway applications: the specification and demonstration of reliability, availability, maintainability and safety (RAMS) part 1: basic requirements and generic process. EN 50126-1. Brussel
10. European Parliament and Council of the European Union (2010) Directive 2010/75/EU on the industrial emissions (integrated pollution prevention and control)
11. Gupta Y, Chow WS (1985) Twenty-five years of life cycle costing-theory and application: a survey. Int J Qual Reliab Manage 2:551–576
12. Harper CL (2004) Environment and society: human perspectives on environmental issues, Taylor and Francis, USA. ISBN-13: 978-0-205-82053-5
13. Kececioglu D (1984) Reliability Educations: a historical perspective, IEEE Transactions on Reliability, vol R-33, No. 1, pp 21–28
14. Kececioglu D (2002) Reliability Engineering Handbook, Volume 1, Pennsylvania. ISBN: 1-932078-00-2
15. Luzzi R, Maida L, Martinetti A, Patrucco M (2013) Information. Formation and training for the maintenance operations: the lesson learned from fatal accidents, chemical engineering transactions 32:229–234. doi:10.3303/CET1332039
16. Markeset T, Kumar U (2003) Integration of RAMS and risk analysis in product design & development: a case study. J Qual Maint Eng, vol 9, Issue 4. doi:10.1108/13552510310503240
17. Mitchell JS (2002) Physical asset management handbook, 3rd edn. Clarion Technical Publishers, Houston, TX
18. Mulder W, Blok J, Hoekstra S, Kokkeler FGM (2012) Design for maintenance. Guidelines to enhance maintainability, reliability and supportability of industrial products. Enschede, University of Twente. ISBN: 978-94-6190-993-0

19. Parada Puig JE, Basten RJI, van Dongen LAM (2015) Understanding maintenance decisions: how to support acquisition of capital assets. In Through-life engineering services decision engineering, Springer International Publishing Switzerland. doi:10.1007/978-3-319-12111-6_13
20. Patton J. (1980) Maintainability and maintenance management, Research Triangle Park, N.C: Instrument Society of America
21. Rajabalinejad M, Martinetti A, van Dongen LAM (2016) Operation, safety and human factors: critical for the success of railway transportation. In Systems of systems engineering, Kongsberg, Norway, 12–16 May 2016
22. Ruitenburg RJ, Braaksma AJJ, van Dongen LAM (2017) Asset life cycle plans: twelve steps to assist strategic decision-making in asset life cycle management. Optimum Decision Making in Asset Management, IGI Global
23. Todinov MT (2005) Reliability and risk models: setting reliability requirements. John Wiley & Sons Publication. doi:10.1002/0470094907.ch1
24. Umeda Y, Takata S, Kimura F, Tomiyama T, Sutherland JW, Kara S, Herrmann C, Duflou JR (2012) Toward integrated product and process life cycle planning an environmental perspective. CIRP Ann Manuf Technol 61(2):681–702. doi:10.1016/j.cirp.2012.05.004
25. United Nations Environment Porgramme UNEP (2011) Towards a green economy: pathways to sustainable development and poverty eradication. www.unep.org/greeneconomy
26. van Dongen LAM (2015) Through-life engineering services: the nedtrain case
27. World Health Organization (1946) Preamble to the constitution of the world health organization as adopted by the international health conference, New York, 19–22 June 1946; signed on 22 July 1946 by the representatives of 61 states (official records of the world health organization, no 2, p 100) and entered into force on 7 April 1948

Chapter 26
On the Initial Spare Parts Assortment for Capital Assets: A Structured Approach Aiding Initial Spare Parts Assortment Decision-Making (SAISAD)

A. Martinetti, A.J.J. Braaksma, J. Ziggers and L.A.M. van Dongen

Abstract In the capital-intensive industry, maintenance expenditures can add up to several times the initial investment. In order to be competitive in their business, owners and users of these capital goods have to take into account the total life cycle cost at investment (e.g. the lifespan of a capital is often more than several decades), the renewal decisions for their installations and the logistic management of the spare parts. Erroneous or unstructured initial spare parts assortment decision-making part of the logistic management can lead to undesired downtime and increases the risk of obsolete or unavailable components. Decision making is complicated by non-existent data in the early design phase and several information management problems. Based on a case study at Netherlands Railways (the largest maintainer of rolling stock in the Netherlands) and literature review a Decision Support Model to structure and improve the data gathering for more effective initial spare part assortment decision making is proposed.

26.1 Introduction

When a new capital asset is introduced there has to be made a timely decision on the initially needed spare parts to have the asset owner provide adequate, safe, reliable capital assets every day. The initial spare parts decision has not only to be made just but also in time because of the lead times. Research shows [8] this lead times can be variable and long and therefore need to be made before construction starts and often

---

A. Martinetti (✉) · A.J.J. Braaksma · J. Ziggers · L.A.M. van Dongen
University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands
e-mail: a.martinetti@utwente.nl

L.A.M. van Dongen
e-mail: l.a.m.vandongen@utwente.nl

L.A.M. van Dongen
Netherlands Railways, Stationshal 15, 3511 CE Utrecht, The Netherlands

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_26

before reliable quantitative data is available. However ordering unneeded or wrong spare parts can lead to a significant increase in the spare part holding costs.

Having the right information at the right moment is however complicated by several challenging issues in data management. Therefore not only a good decision making model is needed but also a decision process which is adaptable to the available information. Having the right information at the right time is a serious challenge because of several data hand-over related problems: (i) fragmented information problems, (ii) high cost for gathering detailed information, (iii) information not easily provided due to strategic reasons of suppliers, (iv) content uncertainty with regard to the provided information and (v) future spare part availability uncertainty.

First of all the present paper will review the data requirements of existing models for spare parts assortment decision making and then will explore the information problems with the application of these models in practise based on a case study at Netherlands Railways. Secondly criteria and principles for the design of a decision support method will be discussed. Finally a decision support method will be proposed to structure initial data gathering and decision making.

26.2 Initial Spare Parts Management

Although there is plenty of literature on spare part management; they all assume the initial spare parts assortment as ‘given’. According to Driessen et al. [4] there is no specific literature which can help to support the decision whether to add a spare part to the initial assortment or not.

However, closer analysis of the literature reveals that some literature sources discuss the (initial) spare parts assortment decision-making process without going much into detail.

Almost all quantitative methods are based on the Recommend Spare Parts List (RSPL), the failure rates λ or on the Mean Time Between Failure (MTBF). These data should be provided by the supplier or is based on operational experience.

The most prominent available quality method is the method of Jones [7] which uses Maintenance Task Analysis (MTA) to determine the spare parts assortment. However, for the initial spare parts decision making there is often not enough data available to make a detailed MTA possible and worthwhile in the initial stage.

Table 26.1 lists some of the identified literature and show the relevant keywords and the data used for the assortment decision making.

However as initial part decisions have to be made on time, the right characteristics have to be known or determined. Therefore the usability of the aforementioned methods is highly information and data dependent.

26.2.1 Spare Parts Characteristics

In order to successfully manage spare part inventories the maintenance department needs to know the different characteristics of every single spare part (or at least of

Table 26.1 Relevant data literature on the most common spare part decision models

Name of the method	Used input data
NEN-EN 13460 [10]	RSPL, MTBF[a]
Driessen et al. [4]	RSPL, MTBF, operational costs
Jones [7]	RSPL, λ, MTA[b]
Diallo et al. [3]	RSPL, λ, FMECA, reliability, availability[c]
Aït-Kadi et al. [1]	RSPL, criteria based on available data
Smit [11]	Similarity, criticality, MTBF, safety, holding costs, lead team, maintenance costs
Keywords used: documentation, maintenance, assortment management, decision, spare parts management, multi-criteria, spare parts, technical documentation	

[a] *RSPL* recommended spare parts list, *MTBF* mean time between failure

[b] *λ* failure rate, *MTA* maintenance task analysis

[c] *FMECA* failure mode, effect and criticality analysis

every spare part class) to classify parts in the decision-making model and to decide their stocking policy.

As for the decision making on the rights spares, the literature provides an extensive list of common procedure for classifying spare parts; some of these procedures are now very briefly explained in order to make easier the comprehension of the next part of the research work.

26.2.1.1 Spare Parts Characteristics Decision Techniques

The following classification can be used as a starting point in the initial spare parts assortment decision making (Table 26.2).

Table 26.2 Relevant data literature on the most common spare part decision models

Name	Keywords	Description
FSN	Fast, slow, non-moving	Technique based on the usage rate of the spare parts; determining fast-moving parts help in deciding which parts need to be stocked close to the asset
SDE	Scarce, difficult and easy to procure	Technique based on procurement lead time. Scarce items are generally imported and require more than 6 months lead time
VED	Vital, essential and desirable	Technique based on the criticality of assortment. A vital part not available has a major effect on the production downtime
ABC	Always better control	Technique based on annual consumption value calculated with the Pareto analysis assigning all the spare parts to three categories
HML	High, medium, low unit price	Technique based on sorting the spare parts according to their unit price

As suggested by Huiskonen [6], forecasting and planning the demand, the procurement lead time, the criticality, the annual usage value and the unit price of the parts for complex capital assets like trains, could produce an unmanageable amount of classes. Therefore a choice on a combination of categories has to be made.

If more than two criteria are needed, multi-attribute classification models could be used [2, 5] assigning with an Analytic Hierarchy Process (AHP) relative priorities or weights to different criteria.

As a consequence, an important challenge for initial spare parts assortment decision making is gathering the proper information and the technical documentation to identify the most essential spare parts and then get the required information so that a part *can* be ordered if required in the future.

26.2.1.2 Spare Parts Characteristics Decision Techniques

Gathering technical information and documentation is vital to set up a correct spare parts management policy. In terms of resources saving, it could be preferable to spend time on collecting and maintaining technical information (criticality, redundancy, commonality, specificity, substitution, life span, position in the configuration and reparability) from the suppliers for spare parts relatively expensive, not easily available and important for controlling supplying risk and operational costs, than gathering technical information for relatively cheap and easily available parts.

Smit [11] presents a possible description of the documentation typology for spare parts. In Fig. 26.1 these characteristics are shown with the cross-references between the documentation.

Unfortunately, in practice even after extensive efforts it is not always possible to gather all presented information for reasons of confidentiality and marketing strategy related issues.

26.2.2 The Case Study: The Introduction of the Sprinter Light Train (SLT)

In our research we will use the introduction of the Sprinter SLT to research the specific context in which initial spare parts decisions are made and to research the specific problems with regard to data management.

The Sprinter Light Train (SLT) series, Fig. 26.2, was introduced in 2009 to replace the outdated Mat ‘64 train series. The SLTs are electrically driven train built by the Bombardier-Siemens consortium. Netherlands Railways needs to maintain 131 SLT’s: 69 four-carriage (SLT-IV) and 62 six-carriage (SLT-VI) trains for over 30 years. Therefore, a spare parts assortment decision-making process had to be defined.

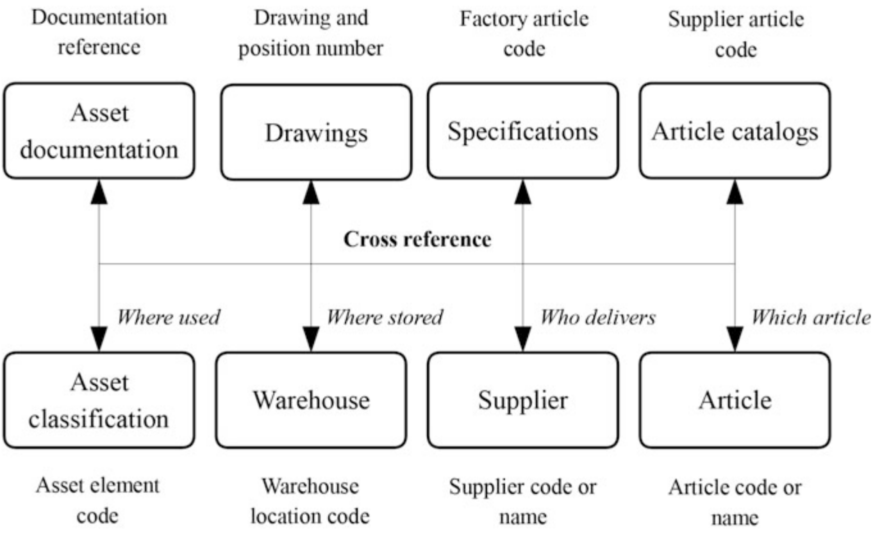


Fig. 26.1 Documentation typology for spare parts [11]

Fig. 26.2 Two Sprinter Light Train's at Netherlands Railways' workshop in Leidschendam, The Netherlands

26.2.2.1 The Sprinter Light Train Assortment Decision Making Process

The first 2 years after the introduction (start of transporting passengers), the supplier is responsible for spare parts management according to the warranty agreement. This means that the supplier has to deal with parts that fail unexpectedly, so-called infant mortality failures.

The Bombardier-Siemens Consortium positioned their engineers and placed their own spare parts in Leidschendam. Whenever a part had to be replaced or repaired because of warranty, the consortium engineers solved the problem.

But, after the warranty period Netherlands Railways takes over this responsibility with the possibility to obtain the spare parts that the Consortium still has on stock.

At Netherlands Railways, the maintenance-engineering department is responsible for the initial spare parts assortment decision making. Staff of this department decides on the assortment with the use of information obtained from suppliers and from own experience. After the initial spare parts assortment is determined, the procurement and logistics staff will decide to purchase parts or not and the amount of spare parts.

Because of Netherlands Railways has to maintain trains for over 30 years, the spare parts assortment is significantly different from the assortment decision made by the supplier for the first 2 years.

Manufacturer delivered spare parts recommendations, failure rates, unit price; minimal order quantity, yearly usage for preventive maintenance information are all potentially available for making the decision. The available information was reviewed by the maintenance engineer (ME) adding information, based on the experience, such as yearly corrective maintenance usage, part classification warranty status, common train series parts and external factors (water, snow and ice effects, vandalism acts etc.).

The initial assortment held about 1600 parts; after few months the assortment was updated to about 2100 parts according to new technical information gathered (part drawings, maintenance manuals and information on form, fit and function.

26.2.2.2 The Data Hand-over Problems and Uncertainties During the Sprinter Light Train Assortment Decision-Making Process

The problems and uncertainties in the spare parts assortment decision-making process of the SLT that arose, were mainly related to the difficult communication with stakeholders causing fragmented technical information and uncertainties about the quality of the gathered data and information:

Fragmented information problems (if Netherlands Railways wants to have more information from the suppliers, the consortium had to be contacted to send the request of information to the suppliers, Fig. 26.3) due the outsourcing [9] of the engineering design, the Netherlands Railways lost full control of the train technology and documentation, this influences lead-times and not always the ‘right’ quality of information is provided;

1. High cost for getting detailed information (gathering all the documentation in a detailed level is expensive causing additional work for the consortium and for the suppliers);

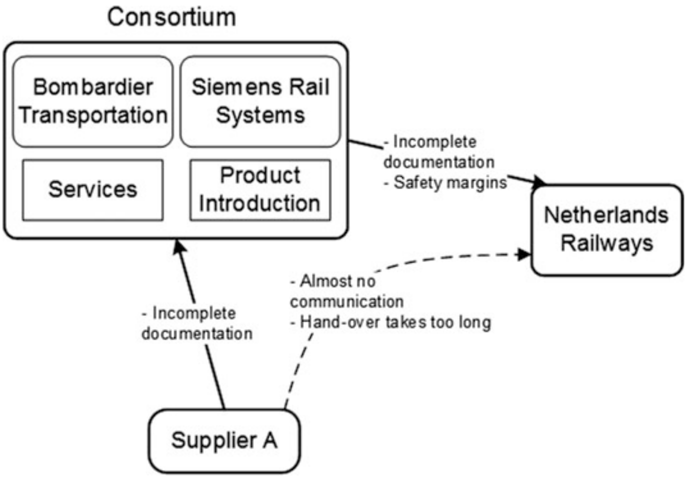


Fig. 26.3 Communication problem between consortium, suppliers and Netherlands Railways

2. Information is not easily provided due to strategic reasons by suppliers because with this information Netherlands Railways can eventually contact suppliers directly to buy parts for lower costs. For the initial assortment Netherlands Railways is however strictly bound to the consortium due to the terms of the contract;
3. Content uncertainty on the obtained information (the safety margins the consortium adds to the safety margins of suppliers cause inaccurate data. Lead times are too long, and parts become more costly than necessary);
4. Future spare part availability uncertainty (the possibility for spare parts included in the assortment to be never used during the lifecycle span of the train) but have to be purchased because there is sometimes only a single option to buy them.

It means that at the beginning of the use phase, the uncertainty on spare parts availability is considerably low; but, according to the lifespan of the train, the uncertainty on future spare parts availability will grow after the introduction.

26.3 Design of the SAISAD Model

In the following sections the SAISAD model will be introduced which is based on solving the aforementioned information management problems by using the earlier defined design criteria and design principles.

26.3.1 Design Criteria and Design Principles

There are some design principles proposed:

1. Gathering information based on asset criticality to reduce communication efforts (problem 1) and costs (problem 2). This also helps to be very focused on the right information at the suppliers, who are not so open in providing asset information (problem 3).
2. Structuring the decision-making process by using a multi-criteria analysis helps to make the process transparent and more fact-driven, thereby reducing cost (problem 2) and prevent issues caused by not including parts in the assortment, considered as the *reverse burden of proof*.
3. Using expert-based information sessions for gathering and selecting the right needed information, where the experts involved can handle content uncertainty (problem 4) and future availability uncertainty (problem 5). Thereby the decision making process becomes robust as it should support the initial spare parts assortment decision making when a relative great uncertainty in the decision phase is present.
4. Using a business case like approach to determine the expected investment and return of the gathered asset information (problem 2).
5. The decision making should be flexible and transparent to allow change when additional information is gathered (problem 4). The spare parts assortment might change a lot after a couple years. For example parts have to be added to the assortment during the using phase due to unexpected failure or difference in the Line Replaceable Unit (LRU)/Shop Replaceable Unit (SRU) level.

26.3.2 The SAISAD Model and Improved Decision Making Approach

To be able to set up a more suitable assortment, a Decision Support Model which combines some spare parts criteria together with a decision making approach is created. The suggested approach was inspired by the method proposed by Aït-Kadi et al. [1].

The SAISAD method, is composed by a logic decision tree approach to create a spare parts assortment, by an evaluation part to consider the Long Term Availability (LTA) and by a special part to take into account the different phase of the process. Hereby an additional step in process has been added (Fig. 26.4).

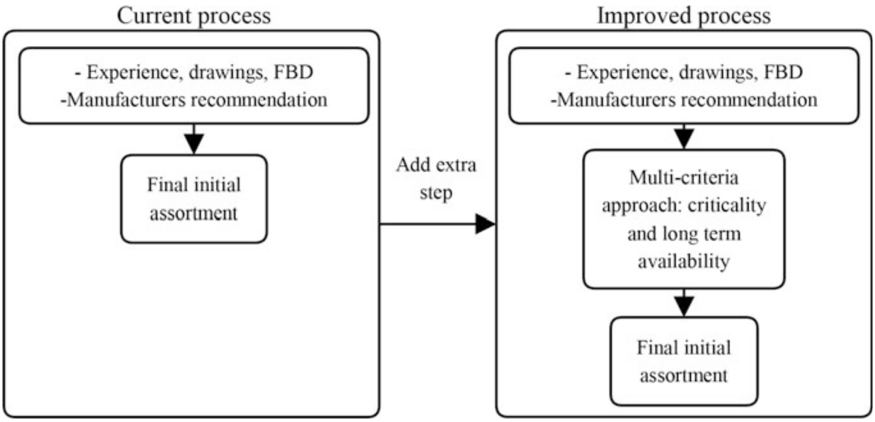


Fig. 26.4 Old approach versus new approach of initial spare parts decision making process

26.3.3 SAISAD Part 1: Logic Tree and Practical Principle

To reduce the number of parts to be analysed in the SAISAD model, a logic tree is used as first filter. To make the filter robust and reliable as much as possible the Pareto principle is considered. The Pareto principle states that in a group a significant 20% of the items contribute to 80% of the total problems and vice versa. Therefore in this case the logic tree should reduce the set of parts to be analysed with about 80%.

The logic tree is based on four decision points (Fig. 26.5):

1. parts that are not sufficiently similar to already known parts;
2. parts that are not decisively influenced by external factors;
3. parts that are significant and have a price higher than 250 €;
4. parts that are not significant and have a price higher than 10.000 €.

The technicians of Netherlands Railways will determine external factors, such as vandalism and risk of fire, during the assortment definition process.

They are not evaluated by the manufacturer because at the initial assortment definition there is no data about vandalism. The decision is then fully based on experience.

26.3.4 SAISAD Part 2: Evaluation Part for the Long Term Availability (LTA)

The evaluation part helps the decision maker to simplify a large complex decision problem into several smaller problems [13], combining Long Term Availability (LTA) of the spare parts identified during the logic tree phase with criticality

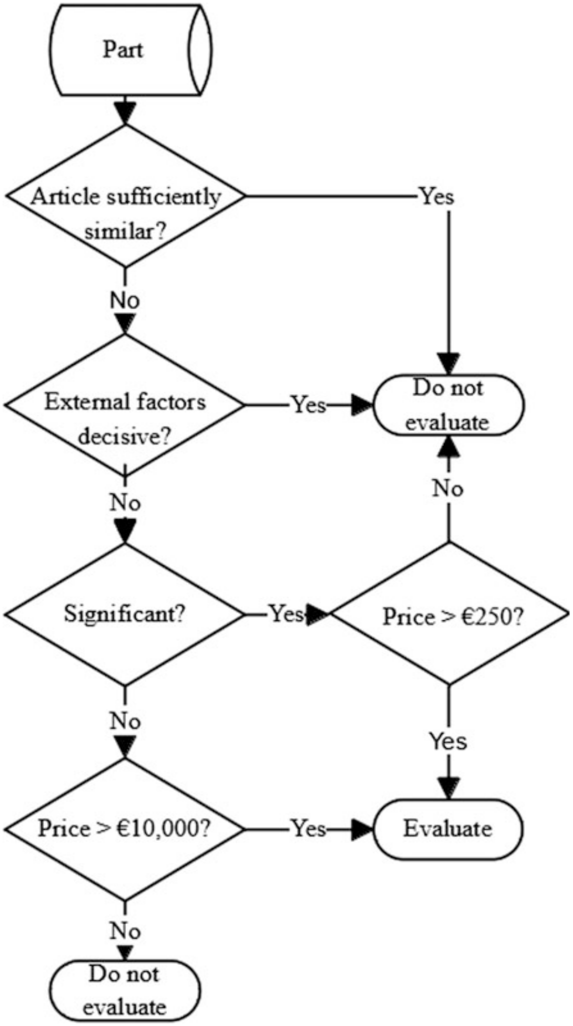


Fig. 26.5 Part 1: logic tree approach of the SAISAD model

categories. The criticality of the part is extracted from the available FMECA documentation related to the part.

The output of the evaluation part is an Importance Factor (1 important—3 least important).

In Fig. 26.6, a Vital criticality and a Vital LTA final importance factor is calculated. The Vital criticality always gives an importance factor 1. When a part is critical, the LTA does not matter: the importance will be high. For parts which have an Essential criticality the LTA does matter; a Vital LTA will yield an importance factor 1. Similarly, Essential and Desirable LTA's will yield an importance factor 2. When the criticality is categorized as Desirable, the LTA can give importance factor 2 (for Vital and Essential LTA) or 3 (Desirable LTA).

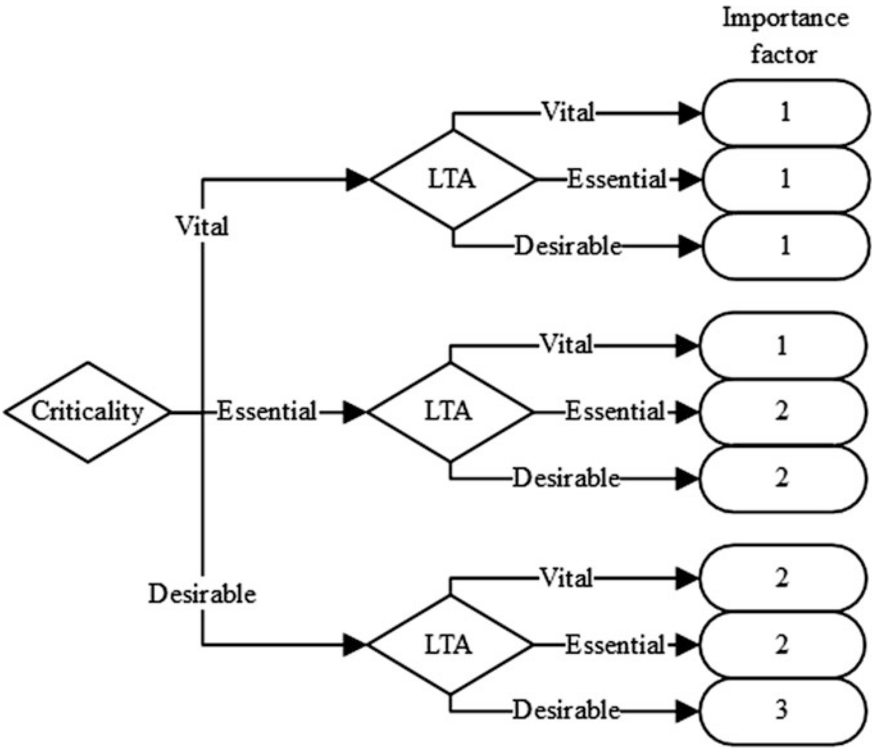


Fig. 26.6 Part 2: long term availability (LTA) combined with vital criticality of spare parts

For each of the important factor decisions, policies are created and assessed in the third part of the SAISAD model.

26.3.5 SAISAD Part 3: Spare Parts Assortment Decision Making

The improved spare parts assortment decision making is formed by an evaluation moment and an expert session.

26.3.5.1 Evaluation Moment

The evaluation moment is created to gather the last information to manage the uncertainty in the initial spare parts assortment decision making. In Fig. 26.7 the advised policy at the first evaluation moment is shown. Parts with importance factor 1 are added to the assortment to be ordered. Parts with importance factor 2 are not

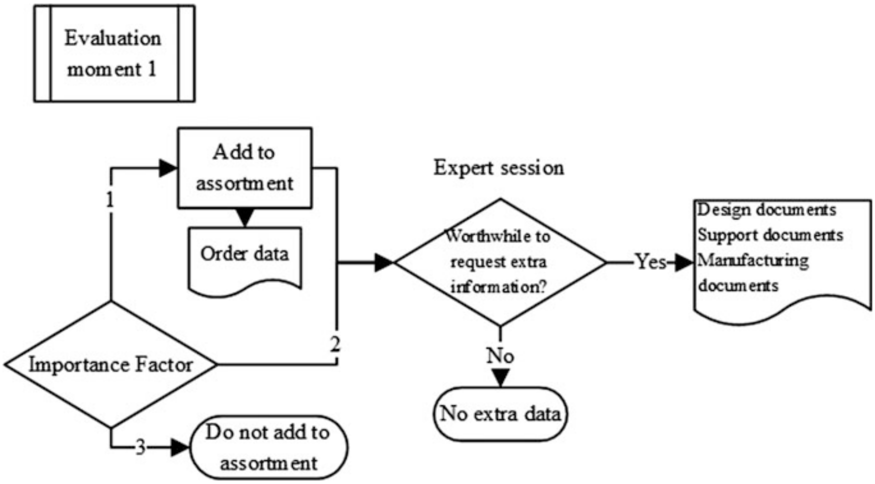


Fig. 26.7 Evaluation moment

added to the assortment at this moment due to the level of uncertainty; the available data is inconclusive on the assortment decision making. Those parts have to be evaluated in an expert session to investigate to require extra information from the supplier and which information has to be requested. Parts with the lowest importance factor (3) are not added to the assortment; no data is gathered because these parts have a low criticality and LTA.

26.3.5.2 Expert Session

To decide on the parts with importance factor 2 the expert session will take place. The maintenance organization needs to have the right technical documentation available reducing and preventing future spare parts unavailability. The people involved in the initial spare parts assortment decision-making process are maintenance engineers and spare parts assortment experts. As discussed in the Sect. 19.2.1 and according to Watts [12], there are several groups of technical documentation could be helpful during an assortment decision making. Due to the complexity of the asset to maintain, the expert session have to take into account different aspects to reduce the uncertainty on the spare parts assortment. The Table 26.3 underlined the documents required during the expert session.

Table 26.3 Technical documentation taken into account during the expert session for the spare parts with importance factor 2

Technical documentation	Description
Sales document	Mostly logistic information about the parts. Lead time, minimum order quantity, storage conditions etc.
Design documents	Design documents define the product or critical process elements. These documents include part drawings, product specifications, parts lists, material specifications, etc.
Support documents	Support documents define the information necessary to install, use or maintain the product. These documents are the product description manual, maintenance manual and installation instructions
Manufacturing documents	Manufacturing documents define the manufacturing process. The tool/fixture drawings, inspection process documents and routing/process sheets. These documents give the needed information to either re-manufacture the parts or to choose external producers to manufacture them

26.4 Conclusions

The design of a structured approach for gathering spare part information in the pre-operational phases is vital to improve the actual spare-part decision making of a typical complex capital asset formed by thousands of parts.

This is hindered by several information management problems which make it necessary to focus efforts. A decision making process is therefore proposed to improve the control on the initial spare parts assortment decision-making process, reducing the number of parts for which data needs to be gathered with a decision tree approach and classifying them according to their criticality and to the Long Term Availability (LTA). The presented SAISAD model is focused on a number of design principles: (1) Gathering information based on asset criticality to reduce communication efforts and costs. (2) Structuring the decision-making process by using a multi-criteria analysis helps to make the process transparent and more fact-driven, thereby reducing cost and prevent issues caused by not including parts in the assortment. (3) Using expert-based information sessions for gathering and selecting the right needed information, with the experts involved can handle content uncertainty and future availability uncertainty. (4) Using a business case like approach to determine the expected investment and return of the gathered asset information. (5) The decision making should be flexible and transparent to allow change when additional information is gathered.

Research Limitations

The LTA is a quantitative estimation of the future availability of the parts and of the supplier affected by an uncertainty. For these reasons, the knowledge of the maintenance experts is of significant value and is acquired during the expert session. The maintenance organization should in addition find ways to decrease the future uncertainty of the spare parts and on the other hand increase the maintainability of the assets ensuring a more dynamic spare parts assortment and long-term availability by optimized contracts with suppliers. Netherlands Railways is therefore planning a new management approach for the introduction of new "Next Generation Trains" on the Netherlands railways. Based on knowledge sharing and cooperation during the planning phase, the aim is to realize a for maintenance and operations designed asset, supported by a reliable spare parts assortment during the life cycle of the asset, delivering the highest value to all involved stakeholders.

Acknowledgements The presented book chapter was inspired by the work of Joost Ziggers at the University of Twente and carried out in collaboration with Netherlands Railways, responsible for the maintenance of rolling stock in the Netherlands.

References

1. Aït-Kadi D, Diallo C, Chelbi A (2003) Spare parts identification and provisioning models. J Decis Syst 12(1):47–65
2. Braglia M, Grassi A, Montanari R (2004) Multi-attribute classification method for spare parts inventory management. J Qual Maint Eng 10(1):55–65
3. Diallo C, At-Kadi D, Chelbi A (2009) Handbook of maintenance management and engineering. In: Ben-Daya M et al (eds). Springer, pp 191–222. ISBN: 1848824718
4. Driessen MA, Art JJ, van Houtum GJ, Rustenburg WD, Huisman B (2010) Maintenance spare parts planning and control: a framework for control and agenda for future research. BETA working paper, no 325
5. Gajpal PP, Ganesh L, Rajendran C (1994) Criticality analysis of spare parts using the analytic hierarchy process. Int J Prod Econ (35):293–297
6. Huiskonen J (2001) Maintenance spare parts logistics: special characteristics and strategic choices. Int J Prod Econ (71):125–133
7. Jones JV (2006) Integrated logistics support handbook, McGraw-Hill logistics series, 3rd edn. McGraw-Hill Professional, pp 18–16. ISBN: 0071471685
8. Martinetti A, Braaksma AJJ, Ziggers J, van Dongen LAM (2015) Initial spare parts assortment decision making for rolling stock maintenance: a structured approach. ESREDA Brussels, Belgium
9. Murthy DNP, Karim MR, Ahmadi A (2015) Data management in maintenance outsourcing. Reliab Eng Syst Saf 142:100–110
10. NEN-EN 13460 (2009) Maintenance—documentation for maintenance
11. Smit K (2014) Maintenance engineering and managements, 1st edn. pp 330–342. ISBN: 978-90-6592-347-8
12. Watts FB (2012) Engineering documentation control handbook: configuration management and product lifecycle management, 4th edn. William Andrew, pp 5–6, 21–48. ISBN: 978-1-4557-7860-7
13. Winston WL (2003) Operations research: applications and algorithms, 4th edn. Cengage Learning, pp 789–795. ISBN: 0534380581

Chapter 27
The Design of Cost and Availability in Complex Engineering Systems

Duarte Rodrigues, John Erkoyuncu and Andrew Starr

Abstract A lot has been covered in literature about equipment/system availability, through-life engineering services (TES) and cost. However, there are still many challenges remaining in the implementation of TES solutions (e.g. maintenance, training, etc.) to achieve cost and availability targets in complex engineering systems. Industry practitioners seek to improve their ability to predict the evolution of complex engineering systems so that it can embrace proactive behaviour to reduce the through-life cost and increase the level of availability. Contracting for Availability (CfA), a commercial process which seeks to sustain a system or capability at an agreed level of availability, is a good example of how many industry organisations have been implementing TES in their business and experiencing difficulties in performing effective cost and availability estimates, in particular at the early stages of the contracts (e.g. bidding stage), where the most of the costs are committed. This chapter focuses on cost and availability trade-off analysis at the bidding stage of CfA, in the defence context. At this phase there is a need to predict the through-life behaviour of the system (e.g. in a 20/30 year life span), which will be affected by risks and uncertainties that complicate the prediction of the total cost and level of system (e.g. platforms) availability. Two case studies with the UK Ministry of Defence (MoD) enabled the identification of the attributes that impact the cost and availability targets of CfA and the key performance indicators (KPIs) to measure the availability. This led to the development of a framework that gives a major contribution to the decision makers by analysing the required investment in each attribute in order to effectively balance availability and affordability at the bidding stage of CfA. This is a novel research in this context as no

---

D. Rodrigues · J. Erkoyuncu (✉) · A. Starr
EPSRC Centre for Innovative Manufacturing in Through-Life Engineering Services,
Cranfield University, Cranfield, UK
e-mail: j.e.erkoyuncu@cranfield.ac.uk; j.a.erkoyuncu@cranfield.ac.uk

D. Rodrigues
e-mail: d.rodrigues@cranfield.ac.uk

A. Starr
e-mail: a.starr@cranfield.ac.uk

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_27

other similar approach has been found in literature. It can be further implemented in practice as a simulation model using mathematical equations and appropriate software tools such as excel or matlab.

27.1 Introduction

The fast development of technologies, information systems, equipment and commercial processes are making industrial systems much more complex. In sectors such as defence, aerospace and energy, this increase of systems complexity aims to respond more effectively to the current sophisticated and challenging type of procurement, where availability is a critical aspect [1–3]. The availability of complex engineering systems is inherently linked to individual equipment availability. To meet the availability targets, organisations are increasingly adopting Through-life Engineering Services (TES) solutions. TES seek to manage and mitigate issues related to the support of complex systems in order to maximise their level of availability and reduce the life-cycle cost [4]. By managing equipment level issues such as damage, no-fault-found or component obsolescence, TES enable both operators and maintainers to achieve the required levels of system availability in a cost effective manner.

In the defence industry, TES have been increasingly adopted by means of contracting practices such as Contracting for Availability (CfA). This is a commercial process established between the Government and industry, where industry is required to provide the necessary support services to meet availability targets of complex military systems (e.g. equipment, ships, platforms, etc.), for long periods of time (e.g. 20 years). This approach has been adopted by many defence departments such as the American Department of Defense (DoD) and the UK Ministry of Defence (MoD). In the US, CfA is also known as Performance-Based Logistics (PBL) [5]. The increasing preference for this type of support contracts rises from the fact that they are normally fixed price which helps the Government to control its through-life costs. Moreover, CfA can include financial incentives that motivate industry to deliver the highest performance services, ensuring that assets are available on demand, at the specified level, and during the entire duration of the contract. A study conducted by Deloitte Consulting for the US DoD, concluded that CfA/PBL generally saves money to the Government and provides superior availability for large scale systems [6]. This fact actually led the DoD to increase its annual investment by 600% in CfA/PBL over the last 16 years. This has been a growing market with high competiveness, which includes organisations such as: Babcock International, BAE Systems, Rolls Royce, Lockheed Martin, Airbus, Finmeccanica, Boeing, Raytheon and Almaz-Antey. Nonetheless, due to the complexity of the military systems and long term nature of the contracts, build cost and availability estimates at the bidding stage is reported to be very challenging [7]. There are many dynamic and interdependent attributes that impact the cost and availability of the systems and complicate the estimates. Therefore, the challenge consists of assessing the right investment in each of those attributes and considering the interdependencies between them, in order to guarantee the highest availability for the lower cost. This process of assessing the gain-loss relationship between cost and availability is commonly known in literature as trade-off analysis [8]. This chapter presents the results from a strong interaction with practitioners from three different defence contractors and MoD consisting of:

1. Reviewing the current practices in designing cost and availability estimates at the bidding stage of CfA (e.g. trade-off analysis);
2. Identifying the important attributes that drive the cost and availability targets in CfA;
3. Identifying the key performance indicators (KPIs) to measure availability in CfA.

In addition, these results underpinned the development of an innovative framework that proposes an iterative process to improve the current trade-off analysis between cost and availability at the bidding stage of CfA. Next chapter describes the methodology adopted to develop this research.

27.2 Research Methodology

This research had a strong interaction with practitioners from three different defence contractors, and MoD, with extensive experience in CfA, in more than 30 h of structured and semi-structured interviews, brainstorming sessions and one workshop. In a total, 17 sessions were performed according to the methodology described in Fig. 27.1.

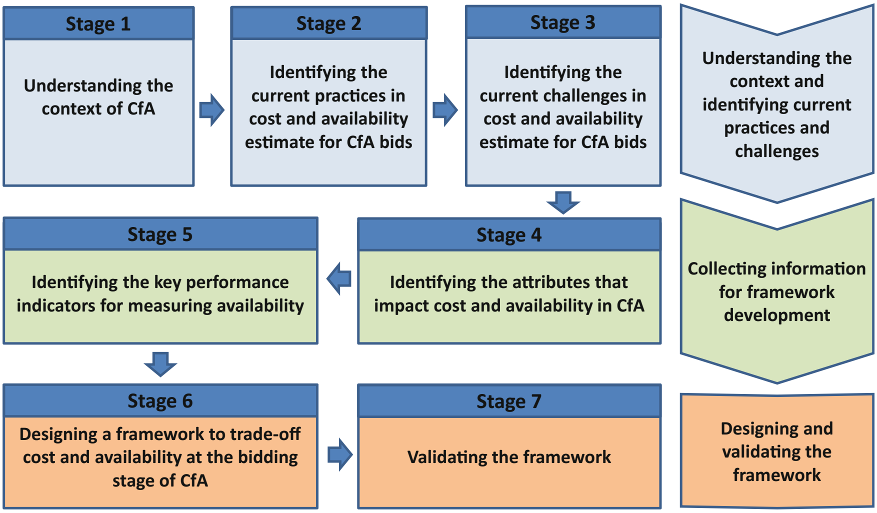


Fig. 27.1 Research methodology process

The people interviewed had an average of 15 years of experience in the context of defence and service-support contracts, and were from different areas of expertise such as modelling, engineering and management, as described in Table 27.2 (see Appendix). The last column of the table indicates the stage(s) of the research where each individual had an impact. In addition, an on-line survey was delivered to 13 senior managers from MoD, defence contractors, and defence consultancy organisations. The respondents had an average of 22 years of experience in defence and roles of management and business development. This multidisciplinary and systematic approach enabled to construct an evidence-based report of the state-of-practice in building CfA bids estimates, identify the current challenges, and develop and validate a framework to improve the current practices by presenting a solution approach to the identified challenges.

27.3 Stage 1: Understanding the Context of CfA

From the extensive interaction with the industry and MoD experts, a definition of CfA has been built, reflecting the perspective of those facing daily challenges in the design of support solutions and estimating for CfA bids: CfA can vary in terms of duration, application, equipment ownership and KPIs. Although, at a high level they all aim at: “making sure that specific equipment will be available to perform a task”. Moreover, the definition of availability is not standardized. Depending on the context of application, there are other alternative and closely related terms that are frequently used such as: “the supply of spares at the right time”, “the supply of specific equipment to be accountable to do a job”, “the management of equipment supplied that has to work when needed” or even “the management of spares on behalf of the costumer”. Such alternatives can cause a misunderstanding of the availability definition and can drive ambiguous requirements and specifications between different types of CfA. Nonetheless, these variations are limited to the boundaries defined in the support options matrix (SOM) model presented in the MoD acquisition operating framework (AOF) [9]. SOM presents the different types of service contracts in the UK Defence and describes the specifications of each contract type. CfA are often fixed price contracts. Therefore, it is of interest for the contractor to optimise the support performance through the contract life-cycle in order to increase profit. For example, repairing the parts instead of replacing the whole equipment is normally cheaper and increase the margin of profit to the company while still ensure that the required availability is achieved. CfA normally follows the **c**oncept, **a**ssessment, **d**emonstration, **m**anufacture, **i**n-Service, and **d**isposal (CADMID) cycle model defined by the MoD, which structures the life cycle of complex equipment. The design stage (corresponding to the concept and assessment phases) is where the most cost commitments are made and hence it is an important phase of the life cycle. It is also where the contract bids are planned and submitted. An ineffective investment at this stage may incur in big cost slippage at later stages of the program life-cycle.

27.4 Stage 2: Identifying the Current Practices in Cost and Availability Estimate for CfA Bids

A number of interviews with experts 1, 2, 3, 4, 5, 6 and 10 enabled to understand how the current cost and availability estimates are performed to bid for a CfA: There are several independent departments within the companies that work together for a common objective: design a support solution that will ensure a certain level of availability of a system and estimate the total cost of the support activities during the entire duration of the contract. Examples of those departments are: commercial department, project management, engineering, procurement, business development, etc. At the top level are the commercial department and the project management; commercial department is the one that interfaces with the customer and the supply chain and wins projects for the company. The commercial department and project management work very closely together and there even are people that work in both departments in parallel. When a new project launches they develop together a document, which is a statement of what a customer requires for that project and then distributes that document for the other several departments. Once the other departments receive the statement of work document they decide which tasks are their responsibility or which ones are not, and send back to the 'top level' an estimate of the time and cost to accomplish the tasks they will allocate. Project managers merge all the estimates from the different departments and build up the final estimate. This estimate consists of a cost estimate, which is the sum of all partial costs sent by the different departments, and an availability estimate, which is built as a function of the estimated time required per each department to accomplish the tasks. They then will try to compare those estimates with similar projects that have been conducted to see if the values make sense and normally try to put pressure on several departments to reduce their estimates in order to build a better final estimate. Normally, most of the estimates are based on expert opinion and historical data. Even though there are some software tools such as RED CUBE and OPUS 10 that can be used to produce cost estimates but they require a lot of detailed information to be inputted which is normally not available at the bidding stage; they are more appropriated to the in-service phase.

27.5 Stage 3: Identifying the Current Challenges in Cost and Availability Estimate for CfA Bids

The high reliance on the comparison with the cost and availability estimates of similar projects that have run in the past can cause inaccuracies, especially because of the rapid development of the technologies and systems that make it harder to compare between old and new systems. Also, there is a recognised difficulty of breaking down the problems and understand what attributes have an impact on the availability of the systems. Moreover, decision makers face major challenges on assessing the impact of those attributes on the cost and availability targets of CfA in order to effectively balance between availability and affordability at the bidding stage. They need to be aided by a structured framework that is able to identify, for each project, what are the attributes that impact the through-life cost and availability of the system and help them to follow a systematic process to perform an assertive trade-off analysis between cost and availability in order to build reliable estimates. To the best of the authors' knowledge, this type of framework is currently missing in both academia and in industry.

27.6 Stage 4: Identifying the Attributes that Impact Cost and Availability in CfA

A number of interviews with the experts 1, 2, 3, 4, 5, 6 and 9 enabled to identify the main attributes that impact the cost and availability targets in CfA. The main attributes were from the first interview identified as:

- Corrective and preventive maintenance;
- Operational Defects (OPDEF);
- Safety conditions for maintenance;
- Unpredictable in the support requirements;
- Obsolescence management;
- Gain share (e.g. performance incentives included in the contracts);
- Training (e.g. CONDO);
- Updating of MoD data system (e.g. MoD Inventory Management Systems (CRISP));
- Asset management.

In the following interviews, the UK defence lines of development (DLoDs) were pointed out by the different experts as an important set of factors that impact the cost and availability of the military support activities. There are eight DLoDs defined by the UK MoD: training, equipment, personnel, information, concepts and doctrine, organisation, infrastructure and logistics. From this stage of the research, the different attributes identified in the first interview started to be seen as a break down structure of each DLoD. However, the list of attributes has not been considered exhaustive yet. Other elements have been further recognised such as the integrated logistic support elements (ILS) [10] and the human factors (HF) [11]. This list has been validated through 10 interviews before the on-line survey described in Sect. 27.3 has been launched. The survey contained a default list of attributes for each DLoD and the respondents selected those that they considered to have an (higher) impact on the effectiveness of each DLoD; they also suggested other attributes that they considered relevant. The final list of attributes included only those attributes that have been selected by at least 50% of the respondents. The finalised list was also validated with experts 1, 2 and 3, which proposed some adjustment based on their experience. As a result, a final list of cost and availability impact factors for CfA was built and is presented in Table 27.1.

Table 27.1 List of attributes that impact the cost and availability targets in contracting for availability

Training attributes		
• Course continuous monitoring and development • Training facilities • Whole-life costs • Manpower • Reliability and maintainability	• Training and training equipment • Training aids and documentation • Personnel • Training development • Maintenance planning • Health hazards	• Training administration • Training personnel • Training needs analysis • Human factors engineering • System safety • Interoperability
Equipment attributes		
• Interoperability • Post design services • Technical information • Maintenance planning • Equipment purchase • Repair and overhaul • Packaging, handling, storage and transportation • Training and training equipment	• Installation and setting to work • Engineering support • Software support • In-service monitoring • Inventory costs • Obsolescence • Health hazard • Equipment support facilities • Whole-life costs	• Configuration management • Technology refresh • Project management • Human factors engineering • System safety • Obsolescence contingency • Reliability and maintainability • Disposal • Equipment
Personnel attributes		
• No of leading hands required • Maintenance planning • Training and training equipment • No of able hands required • Manpower • Whole-life costs	• Health Hazards • Training • System safety • Personnel • Human factors engineering • Interoperability	• Project team required • No of chief procurement officers required • Social and organisation • No of procurement officers required
Information attributes		
• Training • Health hazards • Supply support • Technical information • Maintenance planning • Configuration management • Supporting documentation	• System safety • Plans and associated documentation • Supporting information technology infrastructure • In-service monitoring • Whole-life costs	• Software support • Packaging, handling, storage and transportation (Ils Element) • Reliability and maintainability • Social and organisation • Obsolescence • Interoperability

(continued)

Table 27.1 (continued)

Concepts and doctrine attributes		
• Human factors engineering • Training and training concepts	• Equipment • Manpower	• Personnel • Interoperability
Organisation attributes		
• Interoperability • Social and organisation	• Manpower • Personnel	• Training
Infrastructure attributes		
• Facilities • Whole-life costs	• Interoperability	• Facilities
Logistics attributes		
• Interoperability • Obsolescence • Supply support • Maintenance planning	• Reliability and maintainability • Whole-life costs • Software support • Disposal	• Packaging, handling, storage and transportation • Configuration management • In-service monitoring

27.7 Stage 5: Identifying the Key Performance Indicators for Measuring Availability

The availability is measured as a function of a number of KPIs. To identify what are the KPIs in CfA, an interview with the experts 1, 2 and 3 has been conducted, where they showed a call for bidding document. This document related to a CfA bid for supplying and maintaining a complex medium scale system for 10 years. The KPIs identified were:

- Mean time between failure;
- Ability to support;
- Ability to deploy;
- Ability to recover.

A further interview with experts 12, 13 and 14 enabled to expand this list of KPIs with:

- Reliability;
- Repair Turn Round Time;
- No Fault Found Rates;
- Number of Inventory Spares.

The challenge remains in how to estimate the necessary investment in each TES to achieve the required levels in each KPI during the entire duration of the contract.

27.8 Stages 6 and 7: Designing a Framework to Trade-off Cost and Availability at the Bidding Stage of CfA

This section proposes an innovative framework to tackle the challenges identified in Sect. 27.5. It consists of a systematic process to assess the impact of the attributes identified in Sect. 27.6 in the KPIs identified in Sect. 27.7, towards performing a trade-off analysis between the total cost of the support activities and the level of availability of the system, at the bidding stage of CfA. The process consists of three different phases of assessment as illustrated in Fig. 27.2. Each phase is explained in detail in the next sub-sections.

27.8.1 *Phase 1: Measuring the Cost and Availability Impact of Each Attribute*

As identified in Sect. 27.4, each DLoD has a breakdown structure composed by different attributes. These attributes will directly impact the level of each KPI. Thus,

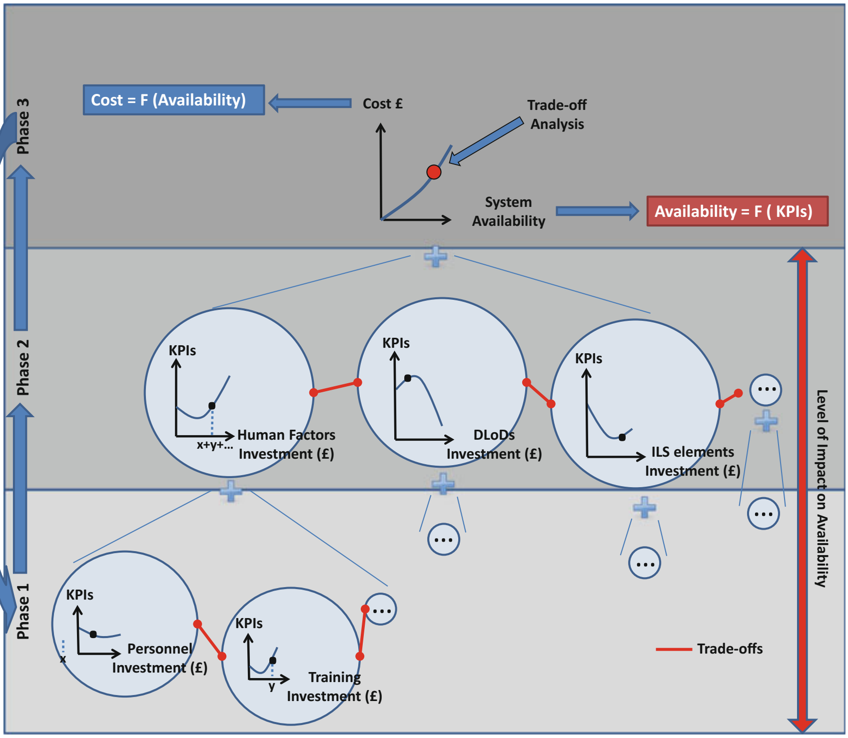


Fig. 27.2 A framework to trade-off between cost and availability in CfA

for each attribute within each DLoD, an effort will be committed towards lifting the performance of each KPI. The sum of the effort in each attribute represents the total effort per DLoD, and the sum of the contribution of each attribute to lift the level of each KPI represents the total impact of each DLoD in the KPIs. The effort in the attributes is considered in the appropriate currency for each attribute (e.g. time, cost, etc.). It is also measured in terms of its cost, being possible to assess the cost of each attribute and the cost of each DLoD. In addition, it considers the impact of each attribute in other attributes. As an example, let us consider the attributes 'training' and 'maintenance', and the currency to measure the effort in those attributes as time. By investing more in training hours there will be, most likely, less need for maintenance hours to achieve the same level of availability. This impact must be considered internally (between the sub-attributes within each DLoD), and externally (between the sub-attributes of the different DLoDs). The assessment of these relationships is considered as a trade-off analysis.

27.8.2 Phase 2: Measuring the Impact of Each DLoD on Availability

The impact of each DLoD on availability is measured through their impact in each KPI. It is calculated by adding up the impact of each sub-attribute in each KPI. At this stage is also possible to identify which DLoD has the highest impact on the performance of each KPI. Trade-offs between the investment in each DLoD can be made by changing the investment in the different (sub)-attributes. The graphs in the figure illustrate the relationships between attributes and KPIs, and the vertical axis on the right had side illustrates different levels of impact of each attribute on availability.

27.8.3 Phase 3: Building the Total Cost and Availability Estimates

At this phase the total cost estimate is simply built by adding up the total cost estimated of each DLoD. Also, the estimated level of availability is calculated by running the availability equation with the estimated level of each KPI. The actual trade-off analysis between cost and availability can now be performed by changing the investment across the different attributes and observing how the total cost and availability change with that. This process can be repeated as many times as needed to get an acceptable balance between cost and availability that guarantees to the contractor a competitive and robust bid estimate. The proposed framework was successfully validated with the experts 1, 2, 3, 7, 8, 9. They said that the proposed framework "reflects exactly what they need" and a practical application of this framework could significantly improve their confidence doing cost and availability estimates for CfA bids.

27.9 Discussion

Decision makers from different defence contractors have in common indicated a need for a framework to help with understanding the key attributes that impact the cost and availability targets in military contracts (e.g. CfA) and to guide their investment across those attributes to achieve a desired performance of each KPI at the bidding stage. Other frameworks presented in literature covering services cost and systems performance have some limitations such as: require too many inputs and a good knowledge about the system to produce estimates [12, 13], do not allow a comparison between operational performance (e.g. availability) and cost performance [14], do not consider the relationships between the cost and performance drivers [15], and focus on a qualitative analysis rather than producing quantitative estimates [16]. The proposed framework distinguishes from the other approaches because it was built upon a detailed analysis of a particular problem offering to the user an exhaustive list of factors that impact the cost and availability performance of the system, minimising the need for inputs and providing insights that enable a wider understanding of the problem. It also focus on producing quantitative and realistic outputs by considering the relationships between the different attributes, rather than just adding up individual contributions. The development and application of this framework to practice can benefit from a number of reviews presented in literature about different techniques to modelling complex engineering systems [7, 17–20]. This framework has recently been used to develop a modelling application [21] which proved to be a valuable aid to the decision makers by giving them more confidence to build cost and availability estimates at the bidding stage of CfA. The same research is currently progressing with developing an optimisation modelling application to perform automatic trade-off analysis between cost and availability in order to find the optimal allocation of a fixed budget across the different attributes, in order to maximise the level of availability during the entire duration of the contracts.

27.10 Conclusions and Future Work

This book chapter presents the results of a structured and iterative research that was planned with the aim of developing the necessary framework. It started by reviewing the definition of CfA and identifying the current practices in performing cost and availability estimates at the bidding stage of CfA. This part was important to design the improvements in the most adequate way.

Thus, the main contributions of this chapter are summarised as:

1. The identification of the current practices and challenges in designing cost and availability estimates for CfA bids;
2. The identification and validation of a list of attributes that impact the cost and availability targets in CfA;
3. The identification and validation of a list of KPIs to measure availability in CfA;
4. A systematic framework to improve the current trade-off analysis between cost and availability at the bidding stage of CfA.

The developed framework enables the decision makers to change their current practices of estimating cost and availability for CfA bids by guiding them at following a more effective process that increases their confidence at balancing affordability and availability.

Further research can now be conducted towards implementing this framework in a practical application by using appropriate mathematical equations and software tools. Also, the future application should be developed in collaboration with the industrial practitioners (e.g. possible end-users of the application) in order to reflect their requirements and limitations (e.g. complexity) in order to increase its usability, utility and acceptance.

Appendix

Table 27.2 Summary of the interaction with defence industry and MoD

Name	Organisation	Job role	Main responsibilities	Experience (years)	Involvement in study
Expert 1	Defence Contractor 1	Project Manager	Design support solutions for military contracts	10	Stages 1–8
Expert 2	Defence Contractor 1	Through Life Support Manager	Design support solutions for military contracts	33	Stages 1–8
Expert 3	Defence Contractor 1	Through Life Support Manager	Design support solutions for military contracts	5	Stages 1–8
Expert 4	Defence Contractor 1	Through Life Support	Design support solutions for military	8	Stages 1–5

(continued)

Table 27.2 (continued)

Name	Organisation	Job role	Main responsibilities	Experience (years)	Involvement in study
Expert 5	Defence Contractor 1	Project Manager	Design support solutions for military contracts	15	Stages 1–5
Expert 6	Defence Contractor 2	Engineering Manager	Policy development and implementation, capability development and management, obsolescence management	25	Stages 1–3 and 5
Expert 7	MoD/Defence Contractor 1	Programme Support Manager	Project planning, engineering design	28	Stage 8
Expert 8	Defence Contractor 1	Modelling Engineer	Create models to support data analysis	12	Stage 4 and 8
Expert 9	MoD	Lecturer in Systems Engineering	Teaching Defence related subjects such as: acquisition, availability definition and sustainability.	20	Stages 1–2 and 8
Expert 10	MoD	Head of Profession for Cost Forecasting within Cost Assurance and Analysis Services (CAAS)	Program management, Operational planning, Cost engineering	15	Stages 3–4
Expert 11	MoD/Defence Contractor 1	Project Manager	Design support solutions for military contracts	15	Stage 8
Expert 12	Defence Contractor 3	Business Development Manager	Developing and managing business solutions	5	Stage 2, 5, and 6

(continued)

Table 27.2 (continued)

Name	Organisation	Job role	Main responsibilities	Experience (years)	Involvement in study
Expert 13	Defence Contractor 3	Project Manager	Business development and contracts design	15	Stage 2, 5, and 6
Expert 14	Defence Contractor 3	Project Manager	Capability development and budget investment management	15	Stage 2, 5, and 6

References

1. Weiss A (2005) Conquering complexity: lessons for defence systems acquisition. The Stationaty Office, London
2. Jones KH, Parker PA, Detweiler KN, Mcgowan AMR, Dress DA, Kimmel WM (2013) Analysis and perspectives from the complex aerospace systems exchange (case)
3. (2013) World energy scenarios: composing energy futures to 2050, London
4. Roy R, Shaw A, Erkoyuncu JA, Redding L (2013) Through-life engineering services. Meas Control (UK) 46:172–175. doi:10.1177/0020294013492283
5. United States Government Accountability Office (2008) Defence logistics: improved analysis and cost needed to evaluate the cost-effectiveness of performance based logistics
6. Deloitte Consulting (2010) Performance based logistics in aerospace and defence
7. Datta PP, Roy R (2010) Cost modelling techniques for availability type service support contracts: a literature review and empirical study. CIRP J Manuf Sci Technol 3(2010):142–157. doi:10.1016/J.Cirpj.07.003
8. Jeang A (2014) Project management for uncertainty with multiple objectives optimisation of time. Cost Reliab Int J Prod Res 53:1503–1526. doi:10.1080/00207543.2014.952792
9. T.N. Archives (2008) Acquisition operating framework. https://www.Gov.Uk/Guidance/Acquisition-Operating-Framework. Accessed 06 June 2016
10. Ministry of Defence (2014) Integrated logistic support Part 1. In: JSP 886 defence support chain management, pp 11–13
11. Ministry of Defence (2015) JSP 912 human factors integration for defence systems Part 2
12. Wendel CH, Braun RJ (2016) Design and techno-economic analysis of high efficiency reversible solid oxide cell systems for distributed energy storage. Appl Energy 172 (2016):118–131. doi:10.1016/J.Apenergy.03.054
13. Murthy DNP, Yeung V (1995) Modelling and analysis of maintenance service contracts. Math Comput Model 22:219–225. doi:10.1016/0895-7177(95)00199-C
14. Bokor Z (2011) Performance-based logistics costing. In: Lindi 2011—3rd IEEE international symposium on logistics and industrial informatics, proceedings, 2011, pp 171–175. doi:10.1109/Lindi.2011.6031141
15. Rahman A, Chattopadhyay G (2008) Cost estimation for maintenance contracts for complex asset/equipment. In: 2008 IEEE international conference of industrial engineering and management, IEEM 2008, pp 1355–1358. doi:10.1109/Ieem.2008.4738091
16. Settanni E, Newnes LB, Thenent NE, Parry G, Goh YM (2014) A through-life costing methodology for use in product–service-systems. Int J Prod Econ 153(2014):161–177. doi:10.1016/J.Ijpe.02.016

17. Curran R, Raghunathan S, Price M (2004) Review of aerospace engineering cost modelling: the genetic causal approach. Prog Aerosp Sci 40(2004):487–534. doi:10.1016/J.Paerosci.10.001
18. Russell R, Chung M, Balk EM (2009) AHRQ technical reviews and summaries, issues challenges conducting systematics reviews to support development of nutrient reference values: workshop summary, US, 2009. http://Www.Ncbi.Nlm.Nih.Gov/Books/Nbk44081/. Accessed 07 June 2016
19. Perkins MJ, Ng TPT, Dudgeon D, Bonebrake TC, Leung KMY (2015) Conserving intertidal habitats: what is the potential of ecological engineering to mitigate impacts of coastal structures? Estuar Coast Shelf Sci 167(2015):504–515. doi:10.1016/J.Ecss.10.033
20. Esteban J, Starr A, Willetts R, Hannah P, Bryanston-Cross P (2005) A review of data fusion models and architectures: towards engineering guidelines. Neural Comput Appl 14:273–281. doi:10.1007/S00521-004-0463-7
21. Rodrigues D, Erkoyuncu J, Starr A, Wilding S, Dibble A, Laity M (2016) A framework to trade-off cost and availability in military contract

Chapter 28
Defence Support Services for the Royal Navy: The Context of Spares Contracts

Davood Sabaei, Alessandro Busachi, John Erkoyuncu, Paul Colegrove and Rajkumar Roy

Abstract "Defence Support Services" (DS2) for the Royal Navy are a particular form of "Product-Service Systems" (PSS). PSS deliver on a turn-key basis equipment/system with related spare parts, training and upgrades to the Royal Navy. In order to stimulate and reward the DS2 provider to improve its services and performance, the Royal Navy wants to shift its contracts from "traditional spare part deliveries" to "performance based contracts" such as "Contracting for Availability" (CfA). However, it has been observed that cost wise, CfA is not logical for all types of complex engineering projects. CfA typically faces higher risks and uncertainties from the solution provider in the early phases of the life cycle (e.g. design phase). There are difficulties in projecting the future costs and required resources in the bidding stage of the contract for the service provider e.g. Obsolescence and required trained workforce for a new introduced technology. These aspects have led some practitioners to prefer spare parts based contracts rather than adopting CfA. However, spare part contracts also have challenges, such as the service provider may not be responsible for the end-to-end process of delivering the support service, and limited time and penalties can cause issues for delivering the service. Moreover, given an extended number of possible solutions, support service providers need an insight from different parts of the supply chain about the cost and time perspectives. This chapter contributes by presenting two novel solutions in spares based contracts including a process to trade-off between time and cost across the supply chain and a framework to assess the costs and benefits of applying "Additive Manufacturing" in the front-end of a DS2 system. Lead time and overall cost are the two main dependent variables across the supply chain. Minimising them lead to a better service delivery. However, there are some challenges for minimising the lead time and overall cost in the supply chain.

D. Sabaei · A. Busachi · J. Erkoyuncu (✉) · P. Colegrove · R. Roy
Cranfield University, Cranfield, Bedfordshire, UK
e-mail: j.e.erkoyuncu@cranfield.ac.uk

L. Redding et al. (eds.), *Advances in Through-life Engineering Services*,
Decision Engineering, DOI 10.1007/978-3-319-49938-3_28

28.1 Introduction to Servitization and Spare Part Contracts

Servitisation is a transformation path for the manufacturing field to develop the capacities to offer and provide services as a supplement to their traditional product offerings. Servitisation is a conventional solution which is growing more in the context of supplement offering services all around the world as well as in the UK [1–3].

Due to market demand, for providing and delivering supplementary and individuals services in different sectors. Several processes and various types of support contracts have been introduced, offered and applied as the solutions for industries. Spare part contracts as part of a servitisation process has been introduced, developed and implemented in the last years and the Royal Navy might be considered as one of the main clients for this type of contracts. For the proposed type of contracts, a service provider delivers support services in order to maintain assets at the desired performance level in the context of the conditions set by the asset owner [4]. The Ministry of Defence (MoD) has introduced three types of spare part contracts regarding their demands. Initial type called “Spares Inclusive Upkeep (SIU)”; in the SIU Industry is contracted for the supply of assets along with the maintenance of spares whilst MOD is still responsible for mending equipment personnel, facilities and technology wise. The second type of spare contracts “Incentivised Upkeep Cost Reduction (IUCR)”; IUCR makes industry responsible for the supply of assets along with the maintenance of spares and is incentivised to reduce the cost of spares. Whereas in the third one which is recognised as Incentivised Reliability Improvement (IRI), industry is contracted for the supply of assets along with the maintenance of spares and is incentivised to reduce costs by improving the reliability of assets [5].

The aim of this chapter is to introduce two solutions, first based on current available technologies and second a futuristic solution for optimising time and cost across the supply chain for delivering spare part contract. To achieve this in Sect. 28.2 a comprehensive explanation for the supply chain for the spare part has been given. Section 28.2.1 describes the first suggested method for optimising time and cost across the supply chain based on current capacities of industries. Section 28.2 is introducing a futuristic method which can have dramatic impact on “Administrative Delay Time” (ADT), “Logistic Delay Time” (LDT) and “Procurement Delay Time” (PDT). Section 28.3 discusses and concludes the chapter.

28.2 The Supply Chain for the Spare Part Contracts

As a general definition, supply chain refers to the flow of raw materials and information from suppliers to end users as a product, as well as the flow of information from the end users to the suppliers [6]. However, supply chain can vary in different contexts. Figures 28.1 and 28.2 illustrate a high level supply chain for delivering a product and a spare part contract as a service.

A general supply chain with high level segments to deliver a product has been extracted from [6] and has been illustrated in Fig. 28.1.

A general supply chain with high level segments to deliver spare part contracts in military context has been extracted from [5] and illustrated in Fig. 28.2.

It can be spotted from Figs. 28.1 and 28.2 that the supply chain for delivering a service specifically delivering a spare part contract has different segments in comparison to the ordinary supply chain for delivering a product.

Studying the supply chain and its segments could be one of the initial endeavours to find ways to reduce the overall cost and to optimise the absolute time for delivering a service or a product regardless of the supply chain type.

Efforts to start identifying relationships between different facilities and segments in a supply chain for an organisation were initiated by [7]. Since then, efforts and endeavours expand on the dependencies between the different segments of the supply chain have been carried on. Studying the required cost and time in different segments and sections of the supply chain can lead to a path to optimise the supply chain overall cost and the lead time.

To clarify the above paragraph, lead time and cost have inverse relation across the supply chain. Often seeking for less delivery time, impose higher cost to service provider. On the other hand, enduring longer delivery time (i.e. to a certain level which does not cause penalising the service provider) slightly reduce costs. Due to above relation between the overall cost and the lead time, there are enormously different answers with different cost and time for delivering a specific service by a certain service provider to the customer. The mentioned logic between cost and time in the supply chain has been illustrated in Fig. 28.3.

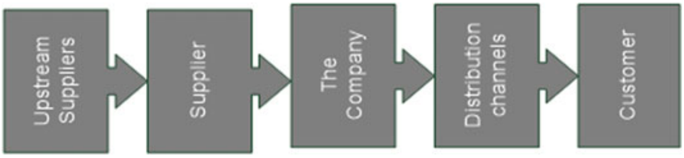


Fig. 28.1 General supply chain for delivering a product

Fig. 28.2 General supply chain for spare part delivery

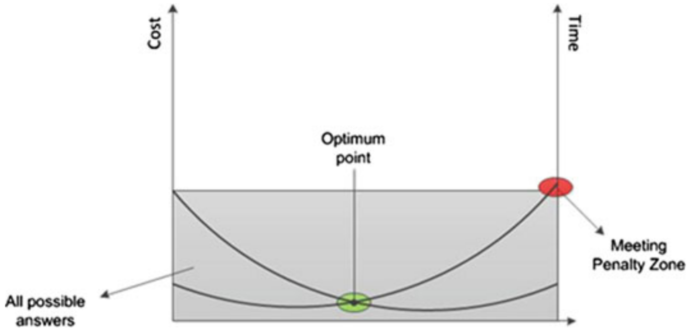


Fig. 28.3 Total cost and time behaviour across the whole supply chain

Selecting a point among all possible answers and then comparing it with all other points can be an approach to help decision makers to take an appropriate decision regarding situations. In addition, among all possible answers there is an optimum point which offers the best service delivery time with minimum cost.

The first solution which is a trading-off tool for helping decision makers has been introduced according to current technologies and considering current capacities. However, second solution which is more a futuristic solution that does not seem costly logical for numerous types of industries nowadays. This futuristic solution is introducing a novel solution to improve spare parts contracts through the exploitation of Additive Manufacturing. Additive manufacturing as a new introduced technology in industries can be used regarding saving time and improving service delivery time.

The first solution provides an optimum delivery time of spares without affecting the cost of delivery of services or it can be used inversely as a framework for finding the optimum point of overall cost of service delivery in a defined time pattern by the customer. The second solution provides an improved delivery time of spares given the elimination of the “Administrative Delay Time” and a dramatic reduction of the “Logistics Delay Time”.

28.2.1 Methodology

The presented framework in the Fig. 28.4 has been developed using “the Soft System Methodology” (SSM) and through primary research based on unstructured interviews with experts of servitisation, service delivery firms and “Ministry of Defence” (MoD). All achieved information was validated by experts in DS2 firms.

One workshop has been held as an initial stage for gathering information and identifying the main perspectives of supply chain from customer and service provider point of view. The main conclusion of the workshop was that lead time and

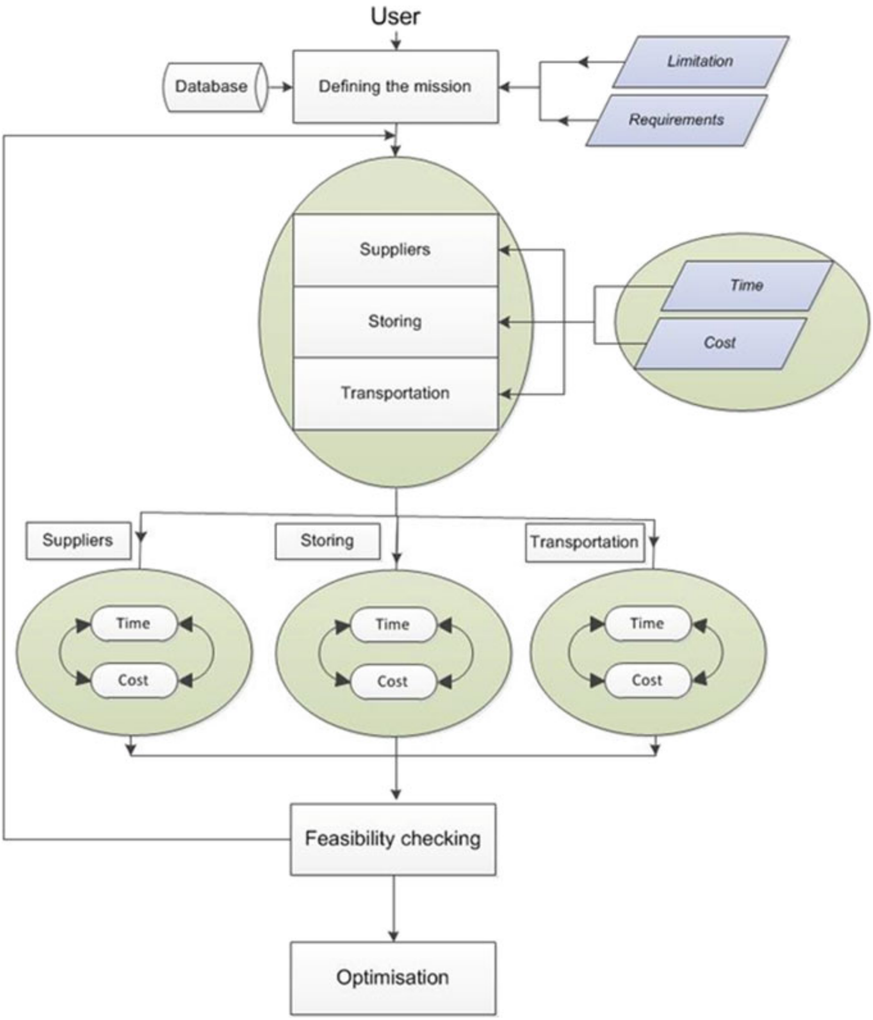


Fig. 28.4 A trade-off analysis framework for delivering spare part contracts

cost as two dependent main drivers across the supply chain and various segments of it, should be the back bone of presented trade-off framework.

Almost 8 interviews have been carried out with individual experts in the servitisation, service delivery and modelling experts. The general outcome of the mentioned interviews was identifying main decision making processes across the supply chain and also categorising the constraints for the optimisation process.

Two interviews have been conducted with contract bidding managers to form the penalising system and its activation process.

The general outcome is the presented framework in Fig. 28.4.

28.2.2 Lead Time and Overall Cost Trade-Off Analysis

As it was discussed in Sect. 28.2, the supply chain of spare part contracts as a part of servitisation process has its own characteristics and segments.

Ability to calculate the effects of modifications in the lead time and the cost in different segments of the supply chain could assist manager for decision making in different aspects e.g. prediction of the consequences of investments in different part of the supply chain, prediction of using or improving alternative methods for manufacturing, transportation and etc. This gap can be filled by presenting a framework which does the trade-off analysis to modify the effect of changes in the lead time and overall cost.

The presented framework in the Fig. 28.4 is a time-cost trade-off analysis which is looking at the supply chain from the service provider's perspective.

According to the defined spare part contracts by MOD which has been fully covered in Sect. 28.1 and also considering the generic spare part contract, which is covered in this section; there are three key decisions which must be taken by service providers. Firstly involved selecting the supplier, second involves selecting a location to store and the method and lastly the transportation method. Each of above mentioned key decisions are based on required delivery time and involved cost of each alternative as the two main drivers.

The introduced framework consists of three main stages. The initial stage is extracting data from the database or generating information about suppliers' details, available transportation systems and their details and storing options and their details as inputs from the user and then by defining the mission the first stage will be finished. User's inputs and identified or extracted information from the database are the main drivers for decision making process. The decision making process is according to trade-off theory. The trade-off occurs among all related costs and time as limitations, requirements and human resources. At this stage trading off occurs in parallel for selecting a supplier, transportation method and storing method.

To clarify the decision making part; time and cost are the two main drivers for trading off and observing the effect of different suppliers, various transportation methods and storing options on the overall cost and absolute lead time of supply chain. In the second stage, a trade-off between two main drivers for various suppliers from different perspective e.g. geographical locations of suppliers and ware houses, required facilities for transportation and storing, the item's cost and required time to deliver a specific item from specific supplier and all other requirements and limitation which were defined in the first stage would be considered.

The third stage in the framework is the feasibility checking step by considering all different cases with different suppliers, storing method and transportation method and effect of each scenario on the overall cost and the total lead time.

There are some constraints for each project which have been defined in the first stage (Defining mission section). The constraints are (1) the maximum delivery time without meeting any penalty for late delivery; (2) defining a method to calculate the penalty system. (3) considering any geographical point for storing. (4) Defining the maximum spendable budget for each section (providing items from suppliers, transportation methods and storing methods) and also defining the maximum spendable budget for the whole project.

All these constraints play a vital role in the optimisation stage. Each of mentioned constraints eliminates a number of possible answers. At the end the optimum point of overall cost within the defined time pattern would be obtained. However, if there is no possible answer in the defined time pattern then the penalising system is going to be activated.

Penalising system is punishing method (or sometimes encouraging method) which often is suggested from customer to service provider. If service provider missed the agreed deadline with a customer then the service provider is going to be under the penalising system. Defining the penalising method often happens in the bid stage of the contract and usually there is an exponential relationship between time and cost while the penalty system is activated.

This section outlined the trade–off framework to help decision makers to find out about different scenarios across the supply chain for delivering spare parts contracts and finding the optimum point for the supply chain. The following section outlines a novel solution to improve spare parts contracts through the exploitation of Additive Manufacturing.

28.2.3 Additive Manufacturing Solution

The following section aims to present the application of "Additive Manufacturing" (AM) within "Defence Support Services" (DS2) and a framework for assessing the impact and supporting the implementation of the AM technology within the Defence sector. The framework is made up of 8 mutually exclusive phases which collectively provide the decision makers an exhaustive analysis of the impact of the AM implementation. As follows, DS2, AM and the framework are presented and explained.

DS2 providers have the capability to deliver the availability of their own or third party systems/equipment to their customer, in this case the Royal Navy. The Royal Navy operates in mission and safety critical environments through the deployment of its platforms. These platforms such as the Type 45 destroyer, Type 23 frigate and the Astute Class submarine are featured with an extended number of sophisticated and complex engineering systems which allows the platforms to deliver its capability and survive in critical and potentially hostile environments. For the Royal Navy the availability of its complex engineering systems is a critical factor which is measured through uptime over total time. The most influential elements of the availability ratio are given by the "Administrative Delay Time" (ADT) and the "Logistic Delay Time" (LDT).

Through the exploitation of AM, DS2 can explore new solutions to support the Royal Navy's complex engineering system. The main idea is to improve the efficiency of the overall service system by eliminating the ADT and LDT through the delocalisation of AM in the front-end of a DS2 system. This solution allows manufacturing the required component next to the point of use. “Additive Manufacturing” (AM) is a disruptive technology which benefits from design freedom, short manufacturing lead times, low buy-to-fly ratios, complexity for free and requires limited space for operating. It can be used for both, printing new components and repair broken ones (if combined with machining and 3D scanner). Moreover the technology has the potential to reduce or eliminate sub-assemblies, access to new geometries and improve the performance of components. AM from a production perspective is lean, it benefits from “pull” and “just-in-time” moreover the technology can process random geometries without any impact on setups. Given the limited space requirements by AM, mini-factories can be developed within containers and deployed in forward bases in order to reduce the distance to the point of use. This allows eliminates the planning of components required (forecasted) and production of only what is actually required in the battlefield. Mini-factories can be developed for in-platform deployment which will eliminate the LDT. Furthermore “Wire + Arc Additive Manufacturing” (WAAM) is an AM technology which is not present in international standards but is considered the most promising technology for industrial applications. Firstly it is a wire based technology which implies no health and safety issues compared with powder solutions, easy material feed, medium cost of wire, nearly 100% material efficiency. Featured with high deposition rates (kg/h), low BTF ratios (2), low cost of investment (max £200 k), high energy efficiency (90%), good accuracy (1–2 mm), low product cost and manufacturing lead times (hrs), the deposition occurs out of the chamber with unlimited size constraints and lower space required. This technology also benefits from good design freedom and topological optimisation opportunity, good mechanical properties and microstructure and no porosity. WAAM is intended for large, fully dense functional components.

To assess the costs and benefits of AM in support services for the Royal Navy and evaluate the impact, a conceptual framework has been developed by [8] and is presented in Fig. 28.5—**Conceptual Framework**. The framework is the result of a collaborative research carried out with the “Ministry of Defence” (MoD), and a leading British Support Service provider of the Royal Navy.

The framework has been developed using “Soft System Methodology” (SSM) and through primary research based on unstructured interviews with experts of DS2 firms and “Ministry of Defence” (MoD). The methodology is outlined in Fig. 28.6 and is made of four phases. Phase 1 consists in the definition of the situation and the problem faced, in this case the emergence of a promising technology, AM and the opportunity to improve the efficiency of the support service system. Phase 2 investigates the current practices, where a system approach has been adopted in order to define a standard of a DS2, its elements, links, triggering events and key performance indicators. Phase 3 involves the development of the framework which is based on the analysis of available AM technologies (from a system perspective) and current DS2 practices. Finally, Phase 4 involves the

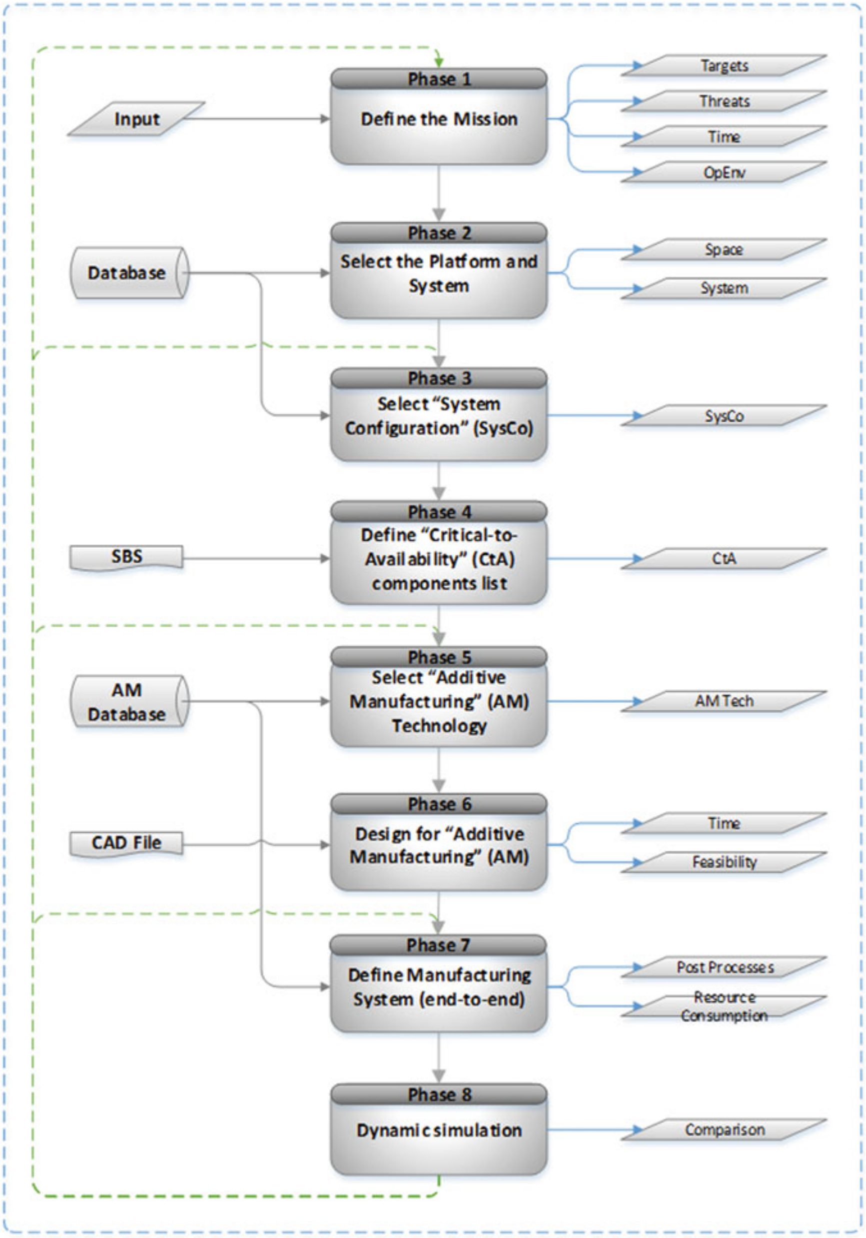


Fig. 28.5 Conceptual Framework

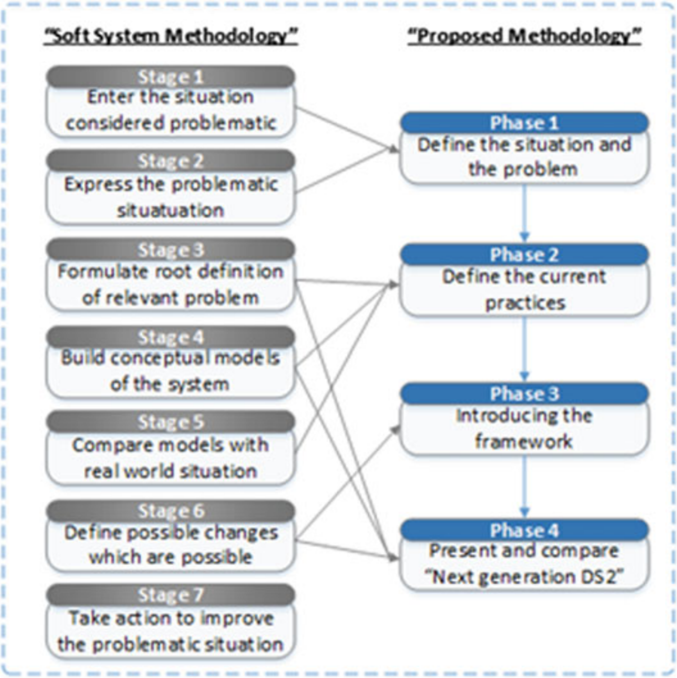


Fig. 28.6 Methodology

comparison of the current practices with the next generation ones based on AM deployed in the front-end of the support service system.

Expertise has been elicited and captured during two workshops which lasted several hours. The results of the workshop have been used to feed a conceptual modelling phase in which the framework has been defined in order to make an exhaustive and holistic assessment of AM applications in DS2. Finally, the result of the conceptual modelling phase outlined in Fig. 28.5—Conceptual Framework have been verified and validated through expert judgement.

The framework has been implemented into a "Decision Support System" (DSS) tool which aims to support critical managerial and technical decision making on the acquisition and implementation of Additive Manufacturing in the Defence Support Service sector. The DSS aims to simulate probabilistically different system configurations available and outline "Key Performance Indicators" (KPI) such as time, cost and benefits. The simulations outlined the following aspects have been observed: (1) AM can be deployed in a defensive platform, a support vessel or a forward base. The impact of AM in DS2 is substantial; firstly it improves dramatically the efficiency of the support to availability of CES, given the elimination of the "Administrative Delay Time" (ADT), "Logistic Delay Time" (LDT) and "Procurement Delay Time" (PDT). Secondly it reduces the supply chain complexity given the supplies of only wire and powders. Thirdly it reduces the time and the cost of the support service with a related reduction of total cost of ownership. Finally, providing flexible manufacturing capability to a defensive platform in a battlefield featured with disrupted supply chain may improve its ability to recover capability and improve its survivability and lethality.

This section has presented a conceptual framework to assess the technical benefits of Additive Manufacturing technology's deployment in the context of support services. Moreover strategic benefits on deployed AM capability have been outlined providing a comprehensive view on the impact of the technology. Results show that AM, if exploited for support service sector, may provide cost, time and availability advantages to both the end user and the service provider.

28.3 Discussion and Conclusion

Due to enthusiasm and demands toward servitisation in different industries such as aerospace and defence, delivering services is becoming more common than before [1]. According to literature, regarding current capacities, extracting and obtaining the maximum output is vital to survive in the current market as well as having futuristic view [9]. This chapter aims to introduce two novel solutions to maximise outputs within the spare parts contract, according to current capacities of industries and a next generation solution.

The first introduced solution is a decision making framework based on current capacities of industries for delivering spare part contracts as part of the servitisation process. Available literature in decision making process for the supply chain in different sector has been studied and the outcome of the literature review was finding a gap in this section. The gap led to develop a framework which has been introduced in Sect. 28.1 of this chapter. The main aim of developed framework is to help decision makers for gaining the optimum decision for delivering services in the spare part contract. Also the framework has an added value for trading off among the enrolled criteria.

The framework is based on trade-off theory consisting of two main role variables (cost and time) across the supply chain. There are three decision making processes involved in the framework, a feasibility checking method and an optimising method at the end. The first decision making is for selecting a supplier, second one is for selecting a transportation method and the third one is for selecting a storing method. Outputs' feasibility of all those three decision making process would be checked in the next stage of framework. All feasible outcomes shall be used in the optimisation stage of framework toward finding the optimum point of overall cost of service delivery in a defined time pattern by the customer. Novelties of the framework consist of 1—Three parallel pairwise decisions making in the service supply chain from providers' point of view and. 2—Considering penalising system and the exponential relation between time and cost while most considered penalising system are linear [10] when the penalty system is activated. The next generation solution, based on the application of AM in the front-end of a DS2 system provides opportunities for dramatic improvement in terms of efficiency and effectiveness of the system. This is mainly due to the reduction of the LDT. Having AM capability in the front-end of the system provides firstly an improved availability of CES and secondly reduces the cost related with delivering the support service. In order to support the acquisition programs on AM and structure the implementation within DS2, an exhaustive framework has been developed. The framework supports critical decision making by providing estimates on costs and benefits of the various AM delocalisation options. The framework has been implemented into a "Decision Support System" (DSS) software tool which allows users to retrieve immediate estimates on AM implementations and compare instantly the KPIs of current and next generation practices. Results shows that the next generation practices based on AM provide improved service and costs savings to both the support service provider and the owner of the platform.

Acknowledgements These researches are performed within the EPSRC Centre for Innovative Manufacturing in Through-Life Engineering Services, grant number EP/1033246/1 and supported by Babcock International.

References

1. Neely A (2007) The servitization of manufacturing: an analysis of global trends. In: 14th European operations management association conference, Ankara, Turkey, pp 1–10
2. Neely A, Benedetinni O, Visnjic I (2011) The servitization of manufacturing: further evidence. In: 18th European operations management association conference, Cambridge, July, pp 1–10
3. Sabaei D, Erkoyuncu J, Roy R (2015) Positioning of spare part contracts in the servitisation process. In: Procedia CIRP, vol 38, pp 106–111. http://www.sciencedirect.com/science/article/pii/S2212827115008859
4. Ng I, Parry G, Wild PJ, McFarlane D, Tasker P (2011) Complex engineering service systems. Springer, London
5. MOD (2007) Contractor logistic support, Defence logistics support chain manual, vol 3
6. Mentzer JT, Keebler JS, Nix NW, Smith CD, Zacharia ZG (2001) Defining supply chain management. J Bus Logist 22(2):1–25
7. Clark AJ, Scarf H (1960) Optimal policies for a multi-echelon inventory problem. Manage Sci 6(4):475–490
8. Busachi A (2016) Additive manufacturing applications in defence support services: current practices and framework for implementation. Flex Serv Manuf J
9. Kastalli IV, Van Looy B (2013) Servitization: disentangling the impact of service business model innovation on manufacturing firm performance. J Oper Manag 31(4):169–180
10. Asgharizadeh E, Murthy DN (2000) Service contracts: a stochastic model. Math Comput Model 31(10–12):11–20